DOCUMENT RESUME

ED 096 155 95 SE 018 102

AUTHOR	Seymour, Lowell A.; And Others
TITLE	Inquiry Role Approach. Field Test Report (1972-73).
INSTITUTION	Mid-Continent Regional Educational Lab., Inc., Kansas City, Mo.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	1 Aug 73
CONTRACT	NE-C-00-3-0068
NOTE	129p.
EDRS PRICE	MF-$0.75 HC-$6.60 PLUS POSTAGE
DESCRIPTORS	Biology; *Curriculum Research; Educational Research; *Inquiry Training; Instruction; Science Education; Science Programs
IDENTIFIERS	Inquiry Role Approach; IRA; McREL; Mid Continent Regional Educational Laboratory; Research Reports

ABSTRACT

Three domains of problems are the subject of this report: problems related to the nature of the Inquiry Role Approach (IRA) project implementation in field test classrooms, problems related to student performance associated with different kinds of extent of project implementation, and problems related to project student performance within the project and as different from comparison groups of nonproject students. Data were gathered by instruments sampling student perception of the existence of certain classroom practices, activity log forms, and monitoring on-site McRel staff. Data on project student performance were gathered by pre-, interim, and posttesting and activity feedback from teachers. Domains measured were: (1) cognitive inquiry process skill, (2) attitude, (3) social skill and small group inquiry activity, and (4) subject matter comprehension. IRA project students showed significant growth in all domains. Comparison nonproject students decreased over a year's time in measured cognitive inquiry skill and attitude while gaining slightly over IRA students in subject matter comprehension. (EB)

ED 096155

U S DEPARTMENT OF HEALTH.
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Lawrence Padberg
McREL
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER"

INQUIRY ROLE APPROACH
FIELD TEST REPORT
(1972-73)

biology
inquiry
role
approach

SE 018 102

INQUIRY ROLE APPROACH

FIELD TEST REPORT

(1972-73)

August 1, 1973

Lowell A. Seymour
Richard M. Bingman
Paul G. Koutnik
Lawrence F. Padberg

Larry L. Havlicek
A. Thel Kocher
Kenneth A. Burton

Do not reproduce in any way without the written permission of...

Mid-continent Regional Educational Laboratory
104 East Independence Avenue
Kansas City, Missouri 64106

The work upon which this publication is based was performed persuant to Contract No. NE-C-00-3-0068, with the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by the National Institute of Education should be inferred.

Special acknowledgment is given to Dr. James Y. Chan for his cooperation in the design and execution phases of the study upon which this report is based; to Mrs. Karen Basey for editing sections of this report and IRA program materials; and to Mr. Michael Hazen for assistance in data processing.

Page

I. SUMMARY 1

II. CONTEXT 2

Program's Beginning (Needs Assessment) 2

History of Inquiry Role Approach (IRA) Program 3

III. PROGRAM DESCRIPTION 8

IV. FIELD TEST 1972-73 15

Statement of Problem 15

Objectives and Null Hypotheses 16

Choosing Participants 21

Description of Participants 24

Procedures for Implementation 32

- Training & Followup 32

On-site Visits 38

Measuring Instruments - Teacher's Log 43
V&P-C 43
CAQ 44
CFE 46
BSBI 46
EIB 1 and 2 48
Social Skills Checklist 52
Attitude Checklist 52
Biology Content Test 53
Understanding Role Responsibilities . 53
DAT 54

Data Analysis & Interpretation 60

Appendix (separate from report) 124

SUMMARY

Pursuant to a contract with the National Institute of Education, the Mid-continent Regional Educational Laboratory has reported on its field test of the Inquiry Role Approach project. The historical context of this field test is described so as to introduce the IRA project itself and give reason for a field test. Three domains of problems are the subject of this report, they are: Problems related to the nature of project implementation in field test classrooms (to what extent was the project -treatment-implemented as planned?); Problems related to student performance associated with different kinds of extents of project implementation; Problems related to project student performance within the project and as different from comparison groups of non-project students.

Data related to extent of project implementation were gathered via instruments sampling student perception of the existence of certain classroom practices, activity-by-activity "log" forms returned by project teachers, and monitoring on-site by McREL staff.

Data on project student performance were gathered via pre, interim, and post testing plus activity-by-activity feedback from project teachers. Pre and post testing was carried out in comparison non-project classrooms. The domains of student performance measured were: 1) Cognitive inquiry process skill, 2) Attitude, 3) Social skill in small group inquiry activity, and 4) Subject matter comprehension.

Specific problems, objectives, and hypotheses which determined the collection of data are described in detail as are the population sampled, the instruments used, the analyses performed and inferences drawn.

The immediate results of data interpretation lead to specific IRA activity materials revisions and suggestions for further study. Specific interpretations and conclusions from the various analyses lead to statements of some confidence that the IRA project was installable to criterion requirements necessary to have it be considered a treatment in this field test. IRA project students grew significantly in cognitive inquiry skills and attitudes over the course of a year's installation, IRA project students performed significantly higher on cognitive inquiry skill and attitude measures than did comparison non-project students. Comparison non-project students decreased over a year's time in measured cognitive inquiry skill and attitude while gaining slightly over IRA project students in subject matter comprehension. These findings are interpreted in light of the characteristics of the instruments used and differences reported between the project and comparison populations.

CONTEXT

PROGRAM'S BEGINNING
(NEEDS ASSESSMENT)

The Inquiry Role Approach (IRA) program was developed during a five-year period (1968-1973) to meet the needs, as seen by some educators, for more participation by students in inquiry activities in which they could assume more responsibility for their own learning. During the school year 1967-68, local Kansas City teachers who were teaching the BSCS* Biology Classes showed considerable interest in using activities that would change the role of students so that the students would become more involved in planning and carrying out investigations for tenth grade biology. During a survey of local high schools, Dr. Richard Bingman, McREL Program Development Specialist, interviewed 10 BSCS Biology teachers and found that generally they desired:

1. An increase in their ability to provide students with more individual attention.
2. More student participation in classroom discussion.
3. Activities that would stimulate student thinking.
4. Better opportunity for students to become involved in the processes necessary to initiate as well as carry out scientific investigations.

These needs corresponded to those expressed by the BSCS staff who developed the three versions of BSCS biology.** During conferences in June and July 1967, concern was expressed that the newly updated modern biology versions (blue, green and yellow) were being taught in a traditional manner and not according to the intent of the developers.*** Based on these needs, the McREL staff focused efforts on designing experiences that would help students assume responsibility for inquiry activities and help teachers acquire the skills and attitudes to assist students. An added incentive to engage in this effort was a lack of inquiry curriculum and teacher education programs at the high school level in science. An analysis of existing inquiry programs showed that nearly all of them were developed for the ninth grade level or were designed for general teacher training. These considerations led to the decision in late 1967 to develop an inquiry program and to concentrate development efforts in high school BSCS biology.

---

* Biological Sciences Curriculum Study, P.O. Box 930, Boulder, Colorado 80302.

** Biological Science: An Inquiry into Life, Yellow Version, Biological Sciences Curriculum Study (BSCS), 2nd edition, Harcourt, Brace & World, Inc., New York, 1968.

High School Biology, Green Version, Biological Sciences Curriculum Study (BSCS), 2nd edition, Rand McNally & Co., Chicago, 1968

Biological Science - Molecules to Man, Blue Version, Biological Sciences Curriculum Study (BSCS), revised edition, Houghton Mifflin Co., Boston, 1968.

*** Summary of Minutes of McREL-BSCS Conference, July 18-19, 1967 Conducted at Boulder, Colorado.

The first such effort was the writing and publication of the document Inquiry Objectives in the Teaching of Biology (IOTB).[1] The publication of this document culminated a joint effort of McREL and BSCS staff aided by prominent science educators to produce a description of desired student outcomes of inquiry activities. This cooperative effort extended over two years and produced a document that has been disseminated extensively throughout the world.[2]

In May 1968 this document was critiqued by a specialist panel of nationally recognized science educators, psychologists and curriculum evaluators.[3] The members of this committee agreed that the attitudinal qualities and cognitive factors presented in the document were important goals for science teaching. Some concern was expressed by individual committee members for certain skills that would enable students to work together more effectively in inquiry activities. Following this session the supplementary document, "Development of Inquiry Skills, Intermediate Objectives"[4] was written to include social skills objectives to facilitate communication, role playing, and conflict resolution skills to improve interaction among individuals or small groups. These objectives along with the attitudinal qualities and cognitive factors mentioned earlier became the basis in 1968 for the development of the Inquiry Role Approach (IRA) and associated evaluation instruments.

HISTORY OF INQUIRY ROLE APPROACH (IRA)

The Inquiry Role Approach (IRA) program has been under development for five years (1968-1973). The stages of development included a Feasibility Study (1968-69), Analyses of Selected Program Elements (1969-70), Formative Evaluation (1970-71), Trial or Pilot Test (1971-72) and Field Test (1972-73). Each of these phases will be described in detail.

Feasibility Study (1968-69)

McREL staff with the cooperation of sixteen local Kansas City BSCS biology teachers prepared and tried out specially designed materials during summer of 1968. These were to be used by teachers and students to initiate role processes and inquiry into the selected concepts and principles which were found in the BSCS textbooks. Pre and post testing was conducted during the school year to determine whether students demonstrated significant gains in acquisition of subject matter concepts, attitudes, and critical thinking skills. An outcome of the testing program,[5] was the occurrance of improved attitude and content acquisition for IRA classes.

This study resulted in a need to develop measurement instruments of the cognitive inquiry factors, which were to be consistent with the definition of inquiry as stated in the IOTB document and operationalized in the IRA program. During the summer of 1969 the development of the Explorations in Biology (EIB) instruments (described in Measuring Instruments Section) was initiated. This instrument was designed to measure students' skill in demonstrating inquiry behaviors in a simulated problem situation. Extremely important to the development of this instrument was the selection of biology topics that would be interesting to biology students. The students in the initial IRA classes cooperated in the selection of these topics.

Based on the 1968-69 study results, local school staff agreed to cooperate with McREL staff to develop the IRA program and a decision was made to move to the second stage of IRA development.

Analyses of Selected Program Elements (1969-70)

During this school year McREL staff focused on a study of the differential impact of using individual work, group work, roles and intermediate testing while working on inquiry guides.[6] The developers believed that increased understanding of the differential impact of these elements would help to establish priorities for what needed to be done during formative development, and in what order. This study was conducted with teachers in Louisiana as well as with a local Kansas City teacher who had participated in the feasibility study conducted the previous year. The local Kansas City teacher participated in a trial test of the materials, procedures and evaluation instruments used later in the Louisiana study. The target population in the Louisiana study consisted of seven BSCS Green Version teachers and nine classes located in seven schools throughout the state. The design of the study included pre, post and intermediate testing either after individual or individual plus group work on special curriculum materials called inquiry guides. The results of the study indicated that while students showed significant gains in acquisition of biological content following both individual and individual plus group work, there were wide variations in student achievement among classes and across inquiry guides. A check on students' work in these classes indicated wide variation in following the suggested procedures for using and discussing the inquiry materials. The work with the Louisiana teachers pointed out an important milestone for development of the Inquiry Role Approach -- the need for an inexpensive and reliable means of determining whether program practices are actually being carried out as prescribed by the developers. This study revealed that unless proper implementation had occurred it was useless to determine its effects. During the 1969-70 study the differential degrees of implementation accounted for much differential impact of the selected program elements. These findings initiated the development of the Views and Preferences, Form A (V&P) instrument (described in measuring instruments section) which is designed to assess students' perceptions of whether selected program practices and activities have occurred and if students have expressed preference for them.

One of the key program practices consisted of the use of small group roles which students carry out as they discuss inquiry problems or work in the laboratory. On the basis of the studies conducted in 1969-70 it seemed that part of the reason teachers did not implement the role practices was due to a lack of clarity of the role behaviors and means of determining when students had acquired sufficient understanding to apply them. On the basis of these findings there appeared to be a need for re-defining the role responsibilities and developing instruments to measure role skill understanding. In the meanwhile, work continued on the EIB, and plans were initiated in 1970-71 for developing a classroom monitoring system.

Formative Evaluation (1970-71)

During the 1970-71 school year intensive development work was carried out with teachers who had considerable experience with the IRA program. They were helped to better define roles, improve the inquiry guides and were given experience in writing manual materials to help other teachers try out the program on a school year basis. The development was focused on writing materials (particularly on roles) and trying them out; revising and writing these materials in a training manual. Because the teachers had had two years of development experience, they felt comfortable enough with the program to try out materials on a day-to-day basis.

During this year another important IRA practice emerged--the use of assessment and evaluation results to improve curriculum materials rather than merely assigning grades--and became incorporated into the manual procedures. As these elements became further developed more attention was given to the sequencing of the activities in which these components; small group, roles and inquiry guides, were used. With the help of the experienced teachers the materials began to form a program.

Trial or Pilot Test (1971-72)

As IRA reached later stages of development, means of implementing it in classrooms outside the local area were developed. The trial test was to provide ample feedback to revise portions or all of the program, as necessary, prior to a more fully expanded field test. In the school year 1971-72 the IRA materials were tested in five high schools of Jefferson Parish, Louisiana with a group of teachers and one supervisor who had not participated in the 1970 study, but who had some orientation to the goals and activities of the IRA program.[7] This testing process enabled the developers to look at the entire program--including the testing program--under the direction of a supervisor and in use by inexperienced teachers.

An important part of the study focused on "proper implementation." The implementation procedures included the classroom teacher practices and the duties of the school district supervisor to coordinate the training and data collection activities. As the supervisor's responsibilities in particular were not successfully carried out, there appeared to be a need to make the program procedures more self-contained and to change the role of the supervisor.

Paralleling the study in Louisiana was a small study in one local school district focused on two experienced IRA teachers who were following the program practices.[8] Two non-IRA teachers in the same building conducted the same testing procedures as the IRA teachers. The results were in favor of the IRA classes and indicated that when the program components were properly utilized favorable results were achieved.

Following the school year 1971-72 test, many revisions were made in IRA materials prior to the subsequent field test. These changes included a redefinition of the trainer role, more directions and options spelled out for teachers (including in-service training being included directly in the manuals), model video and audio tapes used in training, and redesigning of the manuals for easier page and activity location.

During this year the EIB was used for the first time as an evaluation instrument, as was the V&P Form B. On the basis of data, Form B was revised and became V&P Form C and criteria for differential implementation (high, medium and low) were worked out for the field test design to be conducted during the following school year.

Field Test (1972-73)

During the revision process in the summer of 1972 McREL staff decided that the Inquiry Role Approach was readily adaptable to a variety of modern textbook materials in high school biology. Therefore the IRA methodology was adapted to the Yellow Version BSCS textbook.* Previously the methodology had been applied to the Green and Blue Versions of BSCS Biology.* During this revision process the design for the field test for school year 1972-73 was developed and executed.

The 1972-73 field test was undertaken to resolve four problems: Can the adequacy of IRA implementation be described in terms of teacher practices? Do students in classes in which IRA is implemented demonstrate the knowledge and skills which the program materials are designed to develop? Does student performance in IRA classes compare favorably with student performance in non-IRA classes? What recommendations for revision of program materials would be indicated by the field test? These problems and the specific objectives and hypotheses related to each are discussed in a later section of this report.

---

* Biological Science: An Inquiry into Life, Yellow Version, Biological Sciences Curriculum Study (BSCS), 2nd edition, Harcourt, Brace & World, Inc., New York, 1968.

High School Biology, Green Version, Biological Sciences Curriculum Study (BSCS), 2nd edition, Rand McNally & Co., Chicago, 1968.

Biological Science - Molecules to Man, Blue Version, Biological Sciences Curriculum Study (BSCS), revised edition, Houghton Mifflin Co., Boston, 1968.

REFERENCES

1. Bingman, Richard, M., Editor, Anderson, John R., Blankenship, Jacob W., Carter, Jack L., Cleaver, Thomas J., Jones, Willard G., Kennedy, Manert H., Klinckmann, Evelyn, Koutnik, Paul G., Lee, Addison E., Stothart, Jimmy R., Inquiry Objectives in the Teaching of Biology, (a joint publication of Biological Sciences Curriculum Study and Mid-continent Regional Educational Laboratory, September 1969), 147 pages.

2. Bingman, Richard M. "A Report on the Field Use of Inquiry Objectives in the Teaching of Biology" The American Biology Teacher, Vol. 34, No. 4, April, 1972.

3. BSCS-McREL Conference, Kansas City, Missouri, May 12, 1968.

4. "Development of Inquiry Skills Intermediate Objectives," Development of Inquiry Skills staff, Working Paper, 1969.

5. Results described in detail in Learning Through Inquiry, Bingman, Richard M., Neff, Franklin, Havlicek, Larry, Koran, John J., Koutnik, Paul G., McClelland, Nancy, Stothart, Jimmy, 1970. See Part 3, Evaluation: A Progress Report.

6. Results described in detail in "Report on 1970-71 Evaluation and Revision of IRA Component," McREL Staff, 1971.

7. Chan, James, Principal Author. "Report of Field Test, 1971-72 School Year."

8. Results described in detail in "Summary of Local Testing, School Year 1971-72," McREL Staff, 1972.

PROGRAM DESCRIPTION

The Inquiry Role Approach (IRA) is a method of teaching secondary biology which includes teacher training materials, teacher instructions for class use and student materials. While the goals of IRA include the learning of biology content--factual information, concepts and principles of biology--the goals emphasize inquiry skill development, social interaction skill, and attitude development necessary for good inquiry. The IRA method is based on the premise that biology content, understanding inquiry skills, social skills, and attitudes are interdependent and can be achieved best in a program that integrates them. The beginning point and developing rationale for this "four-pronged" approach have been discussed in section II of this report. This section will describe the IRA materials as field tested during the 1972-73 school year.

Thematic Structure of IRA Materials

The materials are divided into three chronological parts called themes. Each theme has a series of activities.* The general pattern of the activities are: 1) An introductory activity to set goals for the themes and, in Themes II and III, a review of the progress made; 2) A series of activities, including laboratory exercises and unique discussion instruments referred to as Inquiry Guides, which develop skills and knowledge needed for the major independent team investigations; 3) An independent team investigation called, Laboratory Explorations in Biology (LEIB) in which the students choose a problem to investigate and then plan, execute, and report the investigation; and 4) Evaluation activities. Each theme has its own emphasis and degree of flexibility and is described below.

In Theme I, students begin an orientation to inquiry by learning the goals of the IRA program, the structure of the four member team, how to use the Inquiry Guide, and how to perform their first major team investigation, LEIB 1. The activities of Theme I, and particularly LEIB 1, are highly structured to help students develop skills, attitudes, and knowledge.

In Theme II structure is decreased. Students now develop and refine their skills using activities more of their own choice than in Theme I. Since student self-direction is influential in determining activities, comprehension and application of social skills, inquiry skills, and attitudes previously introduced are emphasized.

The pre-LEIB activities of Theme III are optional. Students plan, with limited teacher assistance, which activities they will perform chosen from options presented. Each team tailors a set of activities which will provide them with the skills and background information they need for performing their independent team investigation, LEIB 3. The optional activities include exercises in statistical analysis of data and analyses and comparison of science reports. Theme III emphasizes application, analysis, and synthesis of skills and knowledge.

---

* 22 Activities in Theme I, 15 in Theme II, 9 in Theme III.

Thus, IRA presents: 1) A general philosophy for the development of cognitive and social inquiry skills, concepts, and attitudes in biology; 2) A basic teaching method structure consistent with the IRA philosophy; and 3) Activity-by-activity materials which offer decreasing structure and increasing options so that the program can be made specific to the needs of both teacher and students.

Structure of IRA Training Manuals

IRA materials are presented in three manuals. Each manual includes materials for one theme. The manuals are divided into activities. Activities in Theme I are numbered 101 through 122; Theme II, 201 through 215; Theme III, 301 through 309. The decreasing number of activities is consistent with the decreasing structure of the themes. There are four sections within most activities: pre-class teacher instructions, special training, in-class teacher instructions and student materials.

The pre-class instructions section provides a quick overview of the activity for the teacher: 1) A brief rationale for the activity; 2) Specific student objectives expressed in behavioral terms; 3) Time required for the activity--both total time and, if the activity requires several class periods, a more detailed breakdown; and 4) A checklist of the materials and equipment required for performing the activity.

In order to implement IRA, training is necessary for teachers to understand the underlying philosophy, the unique elements of the materials, and how to utilize the methodology to be consistent with the philosophy. Materials for such training have been incorporated into the manuals. The training materials are written so that they can be utilized by a trainer working with one or more teachers or by the teacher who must train himself. These special training materials can be used as a distinct set of training materials in a trainer-directed workshop for several teachers or, because of their placement in the manual, can be used by a self-training teacher. For example, the special trainining designed to explain the overall goals of IRA is found in Activity 103, the same activity in which students are introduced to the goals of IRA. The training designed to explain the structure and use of the inquiry guide--a unique discussion instrument in the IRA program--is found in Activity 107, the activity in which students are introduced to the inquiry guide. Not all activities include a special training section. Most training is found in Theme I (Theme I includes special training in ten activities; Theme II--one activity; Theme III--two activities).

The special training section includes: 1) Rationale for training activity, 2) Objectives for the teacher(s) expressed behaviorally, 3) Time requirements, 4) A checklist of materials and equipment required (video and/or audio tapes are used to model teacher behaviors), 5) Procedures to follow in performing the training (procedures are keyed for use by a trainer with teachers or by a self-trained teacher), 6) Assessment guidelines and criteria for evaluating the assessment, and 7) Any instructional materials required (reading selections, evaluation forms, etc.).

The in-class teacher instructions section provides the teacher with the detailed information required for performing an activity in the classroom. While the pre-class instructions provide an overview, and are generally only 1 or 2 pages in length, the in-class instructions vary from 1 to 21 pages, depending on the requirements of the activity. Organization of the sub-sections of this section vary from one activity to another, depending again on the needs within the activity. All activities in the in-class instruction sections contain assessment guidelines and criteria for evaluating assessment matched to the behavioral objectives stated in the pre-class instructions section.

Student forms are supplied as required for the activity. Twenty-four of the 46 activities include student forms.

Uniqueness of IRA Materials

Small Group Structure - Several factors contributed to the use of four member teams in IRA.* However, the use of teams, or more accurately, groups, is not uncommon. The unique element in IRA small groups is the structure designed to enhance full participation by all team members. Each team member is assigned a role. Each role has a set of related tasks assigned as the role responsibilities of the person who is performing the role. The roles are presented and developed in Theme I first in relation to the completion of a team laboratory exercise; second, in relation to a major team investigation, the LEIB. The four roles are: Team Coordinator--coordinates team discussion, clarifies team direction, summarizes or synthesizes team discussions and decisions; Technical Advisor--assists team in analyzing, challenging, and understanding concepts, principles, statements of evidence, underlying assumptions, etc.; leads team in technical aspects of laboratory work; Data Recorder--records, or directs recording of, data and notes of team discussions; organizes and maintains team records; checks for consistency in records and between records and team decisions or interpretations; and Process Advisor--leads team in analysis of team interaction, identification of strengths and weaknesses, and planning actions to improve teamwork.

Inquiry Guide - The inquiry guide is an instructional device which is designed to organize principles so that the student can see interrelationships. The guides are structured for use of roles in team discussion and for developing higher levels of cognition. Each inquiry guide is structured around a major concept or principle in biology, the major topic. The stem, a brief introductory statement, presents the problem or task, and references are given to information which may be helpful in completing the problem. Five (or more) inquiry statements follow. Each statement presents factual information, an application, an interpretation, an analysis, or some other expression of a sub-principle of the overall principle expressed in the major topic. The student reads the statement, interprets it, states a position (true or false), supports the position with evidence, and records an overall statement of a principle or generalization (the student's understanding of the sub-principle expressed in the statement). Each statement is written in such a way as to require the student to demonstrate a certain level of cognition. As a student works through an inquiry guide, he demonstrates an array of cognitive behaviors.

---

* Discussed in "Learning Through Inquiry," Bingman, et. al., 1970.

The inquiry guide is completed first by all students individually and then in teams. Each student will complete the guide reflecting differences in his experience and understanding. Thus, when the student shares information and ideas during team discussion, new dimensions are added to his understanding of the principles, and he learns about himself and others by having the chance to exchange points of view. The variety of activities engaged in during the team discussion allows distribution of role responsibilities to all members of the team. For example, the Team Coordinator directs discussion and helps the team summarize discussion and synthesize final team decisions. The Technical Advisor helps the team in analyzing guide statements and making interpretations and generalizations. The Data Recorder, not only records team discussions, but watches for logic and consistency in interpretation-position-evidence-generalization. The Process Advisor helps the team review and improve its teamwork.

Laboratory Explorations in Biology (LEIB) - Much of the development of the Inquiry Role Approach has been aimed at operationalizing the inquiry objectives delineated in the BSCS-McREL document IOTB.[1] In this document, six major inquiry factors are identified: 1) problem formulation, 2) hypothesis formulation, 3) design of the study, 4) execution of the design, 5) interpretation of data, and 6) synthesis of knowledge. In most secondary biology classes, students primarily perform steps 4 and 5. The planning phase--(steps 1, 2 and 3)--is presented already completed, and the synthesis of knowledge step 6 is often ignored.

The Laboratory Explorations in Biology are designed to extend student involvement to a full cycle of inquiry. The IRA materials include three LEIB's, one at the end of each theme. The LEIB's, as other elements in the program, become less structured from one theme to the next. LEIB 1 emphasizes problem and hypothesis formulation. LEIB 2 emphasizes design of an experimental study (including use of related literature) and execution of the design. LEIB 3 emphasizes interpretation of data and the application of knowledge gained and synthesis of new knowledge.

In a LEIB, students are presented with a discrepant event, asked what they would question about the event, and are then presented with possible steps to take to answer the question they choose to investigate. Options are allowed for adding or deleting investigatory steps and for using the steps in any order they choose. Some structure is provided to explain each step (what it is, how it can be preformed), but students can always choose whether or not to use a step and the order of use. Completion of the total investigation is followed with a class report by the team, class and teacher evaluation, and replanning of the investigation using the evaluation as feedback.

The LEIB facilitates achievement of the four goals of the IRA program: 1) Inquiry objectives can best be achieved if students are allowed to practice inquiry; 2) As greater responsibility and opportunity for self-direction is given to students, attitudinal development is enhanced; 3) A major team investigation lends itself easily to utilization of the roles and therefore social skill development; and 4) Biological knowledge is sought actively by students rather than received passively since it is sought out and used by students as a necessary part of their investigation. Thus, the LEIB is a culminating activity which efficiently combines all aspects of the IRA program.

REFERENCES

1. Bingman, Richard, M., Editor, Anderson, John R., Blankenship, Jacob W., Carter, Jack L., Cleaver, Thomas J., Jones, Willard G., Kennedy, Manert H., Klinckmann, Evelyn, Koutnik, Paul G., Lee, Addison E., Stothart, Jimmy R. Inquiry Objectives in the Teaching of Biology, (a joint publication of Biological Sciences Curriculum Study and Mid-continent Regional Educational Laboratory, September, 1969), 147 pages.

ADDITIONAL REFERENCES

Andersen, Hans O., Koutnik, Paul G. Toward More Effective Science Instruction in Secondary Education. The MacMillan Company, New York, 1972, pp. 120-125.

Andersen, Hans O., Bingman, Richard M., Humphries, William, Koutnik, Paul G., Lee, Addison, Steiner, H. Edwin, Jr. "Implementing Inquiry Objectives in the Teaching of Biology," School Mathematics and Science, Vol. 71, No. 6, June, 1971.

Bingman, Richard M. Learning Through Inquiry: A Social Design. (Mid-continent Regional Educational Laboratory, Kansas City, Missouri, 1970).

Bingman, Richard M., Koutnik, Paul G. "A Small-Group Study Approach for Biology Based on Inquiry Objectives." The American Biology Teacher, Vol. 32, No. 12, December, 1970.

Bingman, Richard M., Fredrickson, Ronald H., Benson, Barbara. "A Model for Counselor-Student-Teacher Involvement," The Personnel and Guidance Journal, Vol. 50, No. 8, April, 1972.

Bingman, Richard M., Stothart, Jimmy R. "LEIB-IRA: A Preliminary Report," The American Biology Teacher, Vol. 34, No. 9, September, 1972.

Bingman, Richard M. (Contributing author). "What Kind of Environment Will Our Children Have?" AACTE Special Publication, 1971.

Bingman, Richard M. "A Report on the Field Use of Inquiry Objectives in the Teaching of Biology" The American Biology Teacher, Vol. 32, No. 8, November, 1970.

Koutnik, Paul G. "Developing Inquiry Skills: Focus on High School Biology," AETS Newsletter, Vol. 4, No. 1, June, 1970.

Lee, Addison E., Steiner, H. Edwin, Jr. "The Research Potential of Inquiry Objectives," The American Biology Teacher, Vol. 32, No. 9, Dec., 1970.

Lee, Addison E. "Main Points of the McREL-BSCS Document," The American Biology Teacher, Vol. 32, No. 8, November, 1970.

Snowbarger, Barbara C. "New Perspectives for Viewing Inquiry," The American Biology Teacher, Vol. 32, No. 5, May, 1970.

Steiner, H. Edwin, Jr. "A Study of the Relationships Between Teacher Practices and Student Performance of Selected Inquiry Process Behaviors in the Affective Domain in High School Biology Classes," (unpublished doctoral dissertation, The University of Texas at Austin, 1970).

FIELD TEST 1972-73

STATEMENT OF PROBLEM

The 1972-73 field test was undertaken to resolve four problems: Can the adequacy of IRA implementation be described in terms of teacher practices? Do students in classes in which IRA is implemented demonstrate the knowledge and skills which the program materials are designed to develop? Does student performance in IRA classes compare favorably with student performance in non-IRA classes? What recommendations for revision of program materials would be indicated by the field test? These problems are the basis for the specific objectives and hypotheses presented in the following section of this report. The problems, objectives and hypotheses are related as follows:

Problem: Can the adequacy of IRA implementation be described in terms of teacher practices?

Objective 1A

Hypothesis 1A

Objective 1B

Hypothesis 1B

Problem: Do students in classes in which IRA is implemented demonstrate the knowledge and skills which the program materials are designed to develop?

Objective 2

Hypothesis 2

Objective 3A

Hypothesis 3A

Objective 3B

Hypothesis 3B

Objective 4

Hypothesis 4A

Hypothesis 4B

Objective 5

Hypothesis 5

Problem: Does student performance in IRA classes compare favorably with student performance in non-IRA classes?

Objective 6

Hypothesis 6

Problem: What recommendations for revision of program materials would be indicated by the field test?

Objective 7

Hypothesis 7

OBJECTIVES AND NULL HYPOTHESES

Objective 1A - To describe how the program was implemented following each type of teacher training.

Hypothesis 1A - None; this is a descriptive objective.

Objective 1B - To determine whether there is a significant difference in the degree of implementation between groups of teachers receiving different types of training.

Hypothesis 1B - There will be no significant difference between mean implementation ranks (ranking based on the four variables used to describe adequacy of implementation) for teachers receiving different types of training.

Criteria Definitions

Adequate or inadequate implementation of IRA was determined by the extent to which teachers carried out the classroom procedures as stratified by the type of teacher training they received. These two variables are unique to objectives which follow and so they are defined:

Very Adequate Implementation - Three of these four criteria must be met:

1. In Theme I (Teacher's Manual), 90 percent of activities must be completed; Theme II, 70 percent; Theme III, 40 percent.
2. 75 percent of the students must reach the objectives of each activity.
3. Students will respond on the average in the desired way on Views and Preferences C[1] instrument with a mean score of 3.65 or better (views items only).*
4. Students as a group (65 percent or more) agree at the end of Theme II that six of the following nine categories were emphasized as measured by Classroom Activities Questionnaire (CAQ)[2]: application, analysis, synthesis, evaluation, discussion, independence, divergence, ideas valued over grades, and enjoyment of ideas.

---

* Since V&P-C was administered as an interim measure and as a post test the average of these two administrations will be utilized for this hypothesis.

Adequate Implementation - Three of these four minimum criteria must be met. The same criteria definitions are given here as were given in 1 thru 4 above with these changes:

1. Theme I, 80 percent; Theme II, 60 percent; Theme III, 10 percent.
2. 55 percent.
3. A mean score of greater than 3.5.
4. Four of the nine categories emphasized.

Objective 2 - To determine whether there is a relationship between the degree of implementation of the IRA program and student outcomes for biology content achievement, cognitive inquiry skill development, and development of affective qualities of inquiry.

Hypothesis 2 - There is no significant relationship between the three degrees of implementation of the IRA program--very adequate, adequate and inadequate implementation--and the following student outcomes:

1. Comprehensive biology achievement--as measured by the Comprehensive Final Examination-Forms J & K (CFE).[3]
2. Cognitive inquiry skill development--as measured by the Explorations in Biology - Topic 1 (EIB-1).[4] Both the total score and the following sub-scores will be used:
   - A. Formulate a problem
   - B. Formulate a hypothesis
   - C. Design a study
   - D. Interpret data or findings
   - E. Synthesize knowledge gained from the investigation.
3. Affective qualities of inquiry development--as measured by the Biology Student Behavior Inventory (BSBI).[5] Both the total score and the following sub-scores will be used:
   - A. Curiosity
   - B. Openness
   - C. Satisfaction
   - D. Responsibility

Objective 3A - To determine whether IRA students, in classes where the program was at least adequately implemented, will show significant increases in biology content, cognitive inquiry skills and affective qualities of inquiry from the beginning of the school year to the end.

Hypothesis 3A - There is no significant gain from pre- to post-testing in biology content knowledge, cognitive inquiry skills and affective qualities of inquiry--as measured by the instruments described in Objective 2--for IRA students in classes where the program was at least adequately implemented.

Objective 3B - To determine whether IRA students, in classes where the program was at least adequately implemented, prefer the social behaviors, cognitive behaviors and classroom procedures characteristic of the IRA program.

Hypothesis 3B - A majority of the teachers--having performed at least adequate implementation--will not report a class mean score of greater than 3.50 for the preference items on the instrument Views and Preferences - Form C. (A mean score of greater than 3.50 indicates more than 50 percent of the students prefer the set of social behaviors, cognitive behaviors and classroom procedures presented in the instrument.)

Objective 4 - To determine whether there are significant differences in student outcomes in biology content knowledge, cognitive inquiry skills and affective qualities of inquiry between students in the following subgroups:

1. Students in classes where the program was inadequately implemented, adequately implemented, and very adequately implemented.
2. Students with verbal and numerical ability at the 75th percentile or above, from the 50th to the 74th percentile, from the 25th to the 49th percentile, and at the 24th percentile or below.

Hypothesis 4A - There is no significant difference in student outcomes--biology content knowledge, cognitive inquiry skills and affective qualities of inquiry as measured by the instruments described in Objective 2--for students in classes with different degrees of implementation of the IRA program.

Hypothesis 4B - There is no significant difference in student outcomes--biology content knowledge, cognitive inquiry skills and affective qualities of inquiry as measured by the instruments described in Objective 2--for students with different verbal and numerical abilities. (Verbal and numerical abilities: Students will be stratified according to their Differential Aptitude Test[6] pre-test scores into four verbal and four numerical ability groups for tenth grade students only--75th percentile and above, 50th to 74th percentile, 25th to 49th percentile, and 24th percentile and below.)

Objective 5 - To determine whether IRA students will demonstrate criterion level performance in biology content knowledge, cognitive inquiry skills, social skills and affective qualities of inquiry at an interim point in the program.

Hypothesis 5 - IRA students will not demonstrate the following criteria levels when tested at the end of Theme I:

	CRITERION SCORE	PERCENT OF TOTAL SCORE	CHANCE SCORE
1. Theme I biology achievement--as measured by a 72-item biology content test.*			
A. Information and definition items:	12.5	50%	6.25
B. Application and inquiry process items:	23.5	50%	11.75
2. Theme I cognitive inquiry skill development--as measured by EIB-2A & 2B.			
A. Formulate a problem:	1.7	85%	1.6
B. Search for information:	20.9	55%	19.5
C. Formulate a hypothesis:	9.35	55%	7.5
D. Design a study:	28.6	55%	20.4
E. Interpret data or findings:	17.6	55%	16.0
F. Synthesize knowledge gained from the investigation:	9.9	55%	8.6
3. Theme I social skill development--as measured by the Social Skills Checklist* and Understanding Role Responsibilities* quiz.			
A. Understanding Role Responsibilities:	30	75%	10
B. Social Skills Checklist:	28	56%	

---

* These instruments are found in Inquiry Role Approach THEME I MANUAL, Activity 121.

	CRITERION SCORE	PERCENT OF TOTAL SCORE	CHANCE SCORE
4. Theme I affective qualities of inquiry--as measured by the Attitude Checklist.*			
A. Attitude Checklist:	33	51%	

Objective 6 - To determine whether there are significant differences in student outcomes in biology content knowledge, cognitive inquiry skills and affective qualities of inquiry between students in the following subgroups:

1. Students in classes of IRA teachers using the BSCS Yellow Version text.
2. Students in classes of experienced IRA teachers using the BSCS Blue Version text.
3. Students in classes of non-IRA teachers using the BSCS Yellow Version text.

Hypothesis 6 - There is no significant difference in student outcomes--biology content knowledge, cognitive inquiry skills and affective qualities of inquiry as measured by the instruments described in Objective 2--among students grouped by classes of IRA teachers using BSCS Yellow Version, experienced IRA teachers using BSCS Blue Version, or non-IRA teachers using BSCS Yellow Version.

Objective 7 - To determine what revisions in the program materials are indicated by the teacher responses.

Hypothesis 7 - None; this is a descriptive objective. Information for the revision recommendations will be taken primarily from sections 3, 4, 5 and 7 of the Teacher's Log and secondarily from other records of teacher feedback (reports from on-site visits, memoranda and letters from teachers, notes regarding telephone or personal communication with teachers, etc.)

---

* These instruments are found in Inquiry Role Approach THEME I MANUAL, Activity 121.

CHOOSING PARTICIPANTS

During spring 1972, a letter seeking participants for the 1972-73 field test was sent to secondary biology teachers, school administrators, and other educators--college and university personnel, state boards of education personnel, etc. The field test would involve not only classroom teachers, but also trainers of teachers--department chairmen or curriculum supervisors--and, possibly, individuals such as university personnel to train the teacher trainers.

Accompanying the letter was a brief description of IRA and a questionnaire which sought such identifying information as whether or not the person was interested in participating in the field test, in what capacity, and if he could suggest additional persons to contact.*

The initial mailing was sent March 22, 1972 to 47 persons in 16 states. Most individuals were in the McREL region (31 in Kansas, Missouri and Nebraska) and some had had previous involvement with the IRA program.

Lists of secondary biology teachers using the BSCS Yellow Version textbook** in Missouri, Kansas and Nebraska were requested from the respective state departments of education. Partial lists were received and letters were sent to selected teachers (77 in Missouri, 10 in Kansas, 16 in Nebraska) during the month of April. It was found that the lists received were not current. Responses from these mailings were poor, apparently due to the dated information received from the state departments of education. About 10 additional teachers were contacted in various areas as a result of referrals returned to McREL by persons contacted in the initial mailings. Selection of participants from the questionnaire respondents was guided by the following criteria.

---

* See Appendix for letter, questionnaire, and other descriptive materials.

** Biological Science: An Inquiry into Life, Yellow Version, Biological Sciences Curriculum Study (BSCS), 2nd edition, Harcourt, Brace & World, Inc., New York, 1968.

Guidelines for Selection of Field Test Participants

1. A Distribution of Test Sites and a Variety of Trainers - In the 1971-72 pilot test, one school district was used as the test site. All teachers (T_3's) were trained by one trainer (T_2). Any difficulties experienced by the trainer or occurring in the district would affect the complete pilot test sample; therefore several sites and trainers were sought for the field test in 1972-73. A minimum of five trainers was preferred. Since each trainer was expected to train at least 2 teachers, no more than 10 trainers were to be selected since program limitations (number of staff, funding, etc.) would make it difficult to manage a larger number. A minimum of five test sites was preferred. Since teachers without a trainer were also to be included in the field test, the minimum number of test sites was easily achieved.

2. A Variety of Test Site Settings - To avoid bias resulting from a homogeneous test group, a variety of test site settings were sought in regard to socio-economic groups, degree of urbanization, inclusion of ethnic minority groups, etc. Also, in the 1971-72 pilot test students had been segregated according to sex, a factor restricting the variety of students in the IRA classroom and, to a degree, contradictory to the social interaction development fostered in IRA. Such settings were avoided in the 1972-73 field test.

3. Heterogeneity of Student Abilities - While IRA had been formerly operated in heterogeneous classes as well as those with selected ability groups; the intention in IRA development has been to design it for use in heterogeneously mixed classes so that a variety of student skills and backgrounds are brought together. Therefore, only classes heterogeneous in terms of students achievement and abilities were used in the 1972-73 field test.

4. Adequate Sample Size - Krejcie and Morgan[2] have developed a table based on a formula published by the National Education Association[3] for determining sample size in research activities. This table shows that the size of the teacher sample in the field test would not allow for generalization to a large teacher population. For example, a maximum of 40 teachers might be included; results with a sample of 40 can only be generalized to a population of 45. Therefore, our teacher sample size would be determined by other factors--program staff and funding capabilities--rather than generalizability conside considerations.

   On the other hand, Krejcie and Morgan note: "As the population increases the sample size increases at a diminishing rate and remains relatively constant at slightly more than 380 cases." A selection of entries from the table easily demonstrates this:

N (Population size)	S (Sample size)
1000	278
2000	322
5000	357
10000	370
20000	377
30000	379
40000	380
50000	381
75000	382
1000000	384

Therefore, to have the freedom of generalizing to almost any size population of similarly characterized secondary biology students, the student sample in the field test should be no less than 400. This figure was well exceeded.

By June 6, 1972, a tentative field test group had been identified consisting of 6 trainers (T_2's) with 16 teachers (T_3's) plus 7 additional T_3's who would work without trainers. The proposed use of trainers (T_1's) other than McREL staff was dropped. Initially it was thought that T_1's might be utilized to train a number of T_2's in a given area to decrease cost of training all T_2's by McREL staff. At least three persons among educators who had been contacted initially in March 1972 were enthusiastic about taking this T_1-trainer role. Since the tentative list of participants met the guidelines established, further attempts to contact teachers and school personnel for participation in the field test were discontinued. Some difficulties were encountered in regard to administrative support at the proposed test site, conflicting summer commitments, etc. By July 31, 1972, the beginning of the IRA workshop at McREL, the field test participants included: 1) 4 T_2's with 11 T_3's; 2) 4 T_3's without T_2; 3) approximately 1750 students in 65 class sections; and 4) 10 schools in 6 states.

In addition to these participants, eight teachers not using IRA materials, were asked to administer to their classes the battery of evaluation instruments used in the IRA classes. These teachers and their classes were the non-randomly assigned control group; approximately 465 students were included. These students were similar to the test group IRA students in terms of heterogeneous grouping and other factors previously stated. The teachers were also similar to the test group teachers in terms of the textbook they used, experience in teaching, and general teaching approach. The primary difference was the lack of IRA materials and training for the control teachers. Pre- and post-tests were administered in classes of four of the teachers; only post-tests were administered in classes of the remaining four teachers.

DESCRIPTION OF PARTICIPANTS

Of the 19 teachers participating at the beginning of the 1972-73 field test, 4 dropped out (discussed below). The remaining 15 teachers were categorized in 4 groups:

*Group 1: T_2's trained by McREL staff (T_1's); trainers of other teachers (T_3's); teach at least 1 class section.

Group 2: T_3's trained by T_2's.

Group 3: T_3's trained by McREL staff (T_1's) (including 3 T_3's with varying degrees of previous IRA experience).

Group 4: T_2-T_3's (trainer and teachers) working in a team teaching setting.

Category four was necessary because of team teaching taking place at Site B. Student outcomes cannot be related exclusively to any one teacher; therefore, in any statistical analyses by teacher, the five teachers in this group must be treated as one entry.

The participants who completed the entire field test were: 4 T_2's, 7 T_3's with a trainer, 4 T_3's without a trainer, and approximately 1400 students. The eight field test sites, designated A through H, are characterized in Table A.

---

* T_3 = Teacher
T_2 = Trainer of teacher
T_1 = Trainer of trainer

TABLE A: Characterization of Test Sites by Various Descriptors

SITE	TEACHER N	STUDENT N	LOCATION	TYPE OF SETTING	SCHOOL SIZE & TYPE
"A"	1 (T_2)	100	West Coast	Suburban	Large, Sr. H.S., Public
"B"	5 (1:T_2) (4:T_3)	500	Midwest	Suburban	Large, Sr. H.S., Public
"C"	2 (1:T_2) (1:T_3)	210	Midwest	Suburban	Large, Sr. H.S., Public
"D" (2 schools)	3 (1:T_2) (2:T_3)	175	Midwest	Urban	Large, Sr. H.S., Public
"E"	1 (T_3)	150	Midwest	Suburban	Large, Jr. H.S., Public
"F"	1 (T_3)	140	Midwest	Rural-Suburban	Medium, Sr. H.S., Public
"G"	1 (T_3)	80	Midwest	Suburban-Urban	Large, Sr. H.S., Public
"H"	1 (T_3)	30	Midwest	Suburban	Large, Sr. H.S., Public

Criteria for the distribution of test sites, variety of trainers, and variety of test site settings were at least minimally satisfied by this selection of participants.

The four teachers who began the field test but discontinued participation before completion, all expressed difficulty adjusting to the new methodology. Each is discussed individually below.

At field test Site A, the T_2 met only twice with one of the T_3's before the school year began. A second T_3 did not meet with the T_2 and discontinued any efforts to implement the program within the first few weeks. By mid-November the first T_3 also discontinued involvement in the field test, although he continued to use elements of the IRA program materials and administered all evaluation instruments at the end of the year. The T_2 stayed in the program.* At Site C, a T_3 dropped from the program during the two-week workshop directed by the T_2 in late August, 1972. He would have been teaching biology for the first time (he had been teaching earth science) and felt the added demands of learning IRA would be more of a hindrance than a help. A second T_3 at Site C discontinued the program in his class in December.

The T_2 at Site D had a class section with generally low achievement scores because of achievement grouping in other classes within the school. This imbalance made heterogeneous team grouping impossible, and so the T_2 decided to discontinue the field test program in this particular class. He continued the program in another more heterogeneous class section and used parts of the program in his low achievement class.

Teachers

All but 1 of the 15 field test participants had previous teaching experience and previous experience using the BSCS Yellow Version textbook.** The teaching experience of the participants is summarized in Table B.

TABLE B: Years Teaching Experience of Field Test Participants

YEARS EXPERIENCE	NO. OF TEACHERS	TEACHER #
0 - 2	1	11
3-- 5	5	10, 12, 13, 31 & 22
6 - 9	2	02 & 01
10 - 15	3	04, 30 & 14
16 or more	4	40, 20, 21 & 03

Note that the one inexperienced teacher worked in a team teaching setting with four other experienced teachers. Twelve of the fifteen teachers had no previous experience with IRA. Two T_3's at Site F and G had worked with McREL staff during the prototype development of IRA elements (the teacher at Site G during the '68-'79 academic year; the teacher at Site F during '69-'70). One T_3 at Site H had used the IRA materials and acted continuously as a co-developer of materials since the 1968-69 academic year. This teacher was also a co-writer of the IRA materials used in the 1972-73 field test, a member of the McREL staff during the IRA workshop for T_2's and T_3's, summer 1972, and would be categorized more accurately as a T_1- T_3 .

---

* The T_2 designation was kept since this trainer-teacher had functioned in this capacity for almost half of the academic year and continued to have intermittent contact with the T_3 who was still using parts of the program.

** Biological Science: An Inquiry into Life, Yellow Version, Biological Sciences Curriculum Study (BSCS), 2nd edition, Harcourt, Brace & World, Inc., New York, 1968.

Students

IRA is designed for students with abilities and achievement in the 30th to 99th percentile range as measured by the Differential Aptitude Test-Verbal and Numerical[4] (DAT-V&N). Inclusion of students falling below the 30th percentile should not affect the success of the program, neither overall or for those students below the 30th percentile, as long as the student groups are heterogeneous and the percentage of students below the 30th percentile remains low. In the 1971-72 pilot test, student DAT scores (V+N composite) placed students in the following percentile groupings:

Below 30 - 3 percent   50 - 74 - 39 percent
30 - 49 - 26 percent   75 and above - 32 percent

Mean percentile for students in the 1972-73 field test, according to DAT Verbal and Numerical scores, are given in tables C and D.

TABLE C: Mean Percentile for Scores on DAT-Verbal

SITE	N	MEAN RAW SCORE	MEAN PERCENTILE*
A	97	34.27	75
B	508	28.22	57
C	203	29.89	63
D	203	31.77	68
E	141	24.03	43
F	131	27.24	55
G	51	28.47	58
H	19	31.79	68
Total	1318	29.01	60

*10th grade, first semester norms applied.

TABLE D: Mean Percentile for Scores on DAT-Numerical

SITE	N	MEAN RAW SCORE	MEAN PERCENTILE*
A	94	26.56	63
B	456	18.11	30
C	206	23.13	48
D	195	24.48	53
E	124	17.73	27
F	129	21.57	42
G	51	22.51	45
H	18	22.89	47
Total	1240	21.08	40

*10th grade, first semester norms applied.

The DAT-V mean for the entire IRA sample was 29.01, as reported in Table C, and the percentile rank for this mean was 60. The median was 29.00 and the mode was 28.00. Thus these scores were probably normally distributed. The minimum score was 7 and the maximum was 48 on this 50-item test. The standard deviation was 9.94. Thus 84.38 percent of the IRA students had verbal scores at or above the 30th percentile on the DAT-Verbal test.

The mean, median, mode, minimum, maximum, and standard deviations were 21.12, 21.00, 18.00, 1.0, 40.0, and 8.18 respectively for all IRA students on DAT-N. Thus 64.4 percent of the IRA students had numeric scores at or above the 30th percentile on the DAT-Numeric test.

All students were in their first year biology classes using BSCS Yellow Version. Students at Site E were ninth graders; at Site A, students were primarily 11th graders; at all other sites, students were all, or primarily, tenth graders. A large percent of students at Sites B, E, F, and G were below the 30 percentile range. This was higher than preferred.

Characterization of Control Group Participants

All teachers in the control groups were experienced teachers. Classes included were first year biology using BSCS Yellow Version texts composed of all or primarily tenth grade students. Four of the control teachers tested at the beginning and end of the school year; four others tested only at the end of the year. The control sites are characterized by the information in Table E.

TABLE E: Characterization of Control Sites by Various Descriptors.

SITE	TEACHER N	STUDENT N	LOCATION	TYPE OF SETTING	SCHOOL SIZE & TYPE
*A	2	150	West Coast	Suburban	Large, Sr. H.S., Public
C	1	65	Midwest	Suburban	Large, Sr. H.S., Public
E	1	55	Midwest	Suburban	Large, Jr. H.S., Public
H	1	75	Midwest	Suburban	Large, Sr. H.S., Public
I	3	165	Midwest	Suburban	Large, Sr. H.S., Public

* COMMENTS:

A - Same as test Site A; control group was in similar school in same district as test group.
C - This is same school as Test Site C.
E - Same Site as Test Site E; control group in similar school in same district as test group.
H - Same Site as Test Site H; control group in similar school in same district as test group.
I - No test group at this site.

A description of the student populations at the control sites according to DAT scores is given in Tables F and G.

TABLE F: Control Students Percentile Group Distribution According to DAT-Verbal Scores

SITE	STUDENT N	MEAN RAW SCORE	MEAN PERCENTILE*
A	145	35.41	77
C	55	29.78	65
E	51	29.53	63
H	66	30.98	65
I	148	33.10	70

*10th grade, first semester norms applied.

TABLE G: Control Students Percentile Group Distribution According to DAT-Numerical Scores

SITE	STUDENT N	MEAN RAW SCORE	MEAN PERCENTILE*
A	145	29.21	75
C	55	24.98	57
E	51	24.55	55
H	66	24.47	55
I	148	25.95	62

*10th grade, first semester norms applied.

Experienced Teachers

Four teachers in the Kansas City area have participated for five years (1968-69 through 1972-73) in the testing and development of the Inquiry Role Approach program. They are experienced with the BSCS Blue Version* IRA materials. During the 1972-73 school year, they adapted the Yellow Version materials to the Blue Version text and met to exchange ideas among themselves and also give feedback to the McREL staff from their viewpoints as experienced IRA teachers. While not part of the field test group, these teachers provided information and insites beyond the scope of the first year IRA teachers. Evaluation instruments were administered to student samples of each of these teachers. School setting descriptors and student DAT groupings are given in Tables H, I and J.

---

* Biological Science - Molecules to Man, Blue Version, Biological Sciences Curriculum Study (BSCS), revised edition, Houghton Mifflin Co., Boston, 1968.

TABLE H: Characterization of Sites of Experienced IRA Teachers by Various Descriptors

TEACHER CODE	SITE CODE	STUDENT N	TYPE OF LOCATION	TYPE OF SETTING	SCHOOL SIZE & TYPE
61*	H**	90	Midwest	Suburban	Large, Sr. H.S., Public
62	H**	90	Midwest	Suburban	Medium, Sr. H.S., Public
63	J	140	Midwest	Urban	Medium, Sr. H.S., Parochial
64	H**	140	Midwest	Suburban	Large, Sr. H.S., Public

TABLE I: Experienced IRA Teachers' Students Percentile on DAT-Verbal Scores

TEACHER	STUDENT N	MEAN RAW SCORE	MEAN PERCENTILE*
61	17	30.1	63
62	28	32.5	70
63	72	31.8	69
64			

*10th grade, first semester norms applied.

---

* Teacher 61 is also included in the field test group as Teacher #04. This teacher taught one section using the BSCS Yellow Version text (included in the field test group) and three sections using the BSCS Blue Version text (included here).

** Three schools within the same district are represented.

TABLE J: Experienced IRA Teachers' Students Percentile on DAT-Numerical Scores.

TEACHER	STUDENT N	MEAN RAW SCORE	MEAN PERCENTILE*
61	17	23.4	49
62	28	26.7	64
63	72	24.2	53
64			

*10th grade, first semester norms applied.

Students were in first year biology. Students of Teacher 63 were all in ninth grade. Students of the other teachers were all or primarily tenth graders.

PROCEDURES FOR IMPLEMENTATION

Workshops and Follow-Up With Teachers

The implementation plan for the Inquiry Role Approach (IRA) program for school year 1972-73 consisted of three steps:

1. Six participants completed a formal workshop conducted by McREL staff (T_1's). Attending were four trainers-of-teachers (T_2's) who would return to their school districts to conduct similar workshops for other teachers (T_3's) who would in turn implement the program in their classes, and two teachers (T_3's) who represented schools where no one was available to assume the T_2 role.

2. Three T_2's conducted workshops at their schools prior to the beginning of the school year; one T_2 was not able to conduct a formal workshop.

3. Follow-up contacts were made with T_2's and their T_3's, T_1's training T_2's, and T_3's without trainers. Emphasis of these follow-ups was on seeing that each teacher used the Teacher's Manual on a self-contained basis. Discussions among T_2's and T_3's on using these materials enhanced each person's understanding of the IRA process and provided support to the continued commitment to the program. Two T_3's were trained by McREL staff (T_1's) and did not have T_2's or T_3's to share their successes and failures. This study considered important how these T_3's were able to implement the program as compared to T_3's who had assistance. This will be described further in this report.

Workshop Training by McREL Staff (T_1's)

A 10-day workshop was conducted at McREL by four staff members (T_1's) from July 31, to August 11, 1972. The four staff members were developers of the IRA program. (See agenda for the training workshop in Appendix.) The workshop focused on those IRA activities emphasizing small groups, roles, inquiry guides, special laboratory investigations, evaluation procedures, and use of feedback.

The participants were selected as described elsewhere in this report and attended the workshop under the following remuneration agreements:

1. McREL financed room, board, and travel.
2. These provisions did not apply to Test Site A which received only $50.00 for travel.

Each participant in the workshop was encouraged to proceed at his own pace in the in-service activities given in the Teacher's Manual. Each of these activities includes objectives, procedures, reading and assessment materials, and criteria for determining student success. The role of the trainer is specified in the manual under a section called Special Training. The T_1's major function was to model the trainer role as specified in the manual and involved assessing the progress of the participants, applying criteria given in the in-service activity, and providing remedial help when necessary.

Each participant was given the chance to micro-teach certain activities to students brought to McREL. The T_1's coordinated and scheduled these activities with the use of audio and video materials when it became necessary to provide a model for the T_2's to follow on their return to their own workshops. All teachers were given special instructions on how to administer tests and collect and organize data. The T_1's in addition to assessing the progress of the participants, collected data on special problems the participants were having with following manual instructions, time allocations, use of multi-media facilities and materials, and developing implementation plans for their own installations. These data were then used to make final revisions in the Teacher's Manual to be used during the school year.

At the end of the workshop teachers were asked to fill out an evaluation form on the large group sessions, micro-teaching, individualized instruction, and were invited to make other comments on how the McREL workshop was conducted. Generally, the participants favored the format of the general workshop sessions, the micro-teaching, and the individualized sessions of the workshop. Following some remedial activities, all participants met the performance criteria stated for the special training sections of the manual. The remedial activities were required mostly for clarification of procedures, terms, etc., as given in the Teachers' Manual (These have been revised accordingly since the workshop.) Also, suggestions and comments made by these participants helped in replanning the workshop schedule to be useful for subsequent workshops. The replanned workshop schedule differed from the previous one (See Appendix) in these ways:

1. Time for workshop was reduced from 10 to 5 days. In this workshop when teachers were given extra time, they had trouble using it. Also, much time was spent on specific procedural matters pertinent to only one or two members. Such matters should have been discussed during special individual sessions which were scheduled at the end of each day. More rigorous allocation of time was the major change in the workshop schedule.

2. In the new workshop schedule more stress is given to the use of time outside the workshop--preferably in advance of workshop--to carry out reading and planning activities. This change is consistent with one original objective of the IRA program--to make the special training section of the Teacher's Manual as self-contained as possible.

3. Other suggestions include: Use of same students in micro-teaching from beginning to end; More emphasis on video taping of micro-teaching; Use of flow diagrams to present an overview of the IRA program.

The staff noted training difficulties in conducting the workshop included a problem with getting the participants to work individually or, even, in small groups. Even though participants expressed a desire to stay longer in the large group setting, one of their main criticisms in later evaluation focused on--"spending too much time in general or large group sessions." A surprising difference existed among teachers in regard to working independently for reading or planning assignments. Their previous experience at other workshops may have had important effects. Those teachers who had not had previous experience had much more difficulty in adjusting to individualized schedules.

At the end of the workshop the T_2's prepared their own workshop schedules which were examined by McREL staff. Most of these resembled the workshop schedule used during the McREL workshop, except for time allocations. Consideration of the time allocation was due to the fact that the McREL staff had shared their revised workshop plan with T_2's prior to their completion of their own plans. Also, some school districts have restrictions in terms of time allocations. Restrictions on media equipment also required changes in those activities designed for the use of video tapes and 8 mm projectors. Therefore, McREL staff simplified the audio visual requirements for the program or provided alternate activities not requiring such equipment.

Final concluding comments on the workshop in general included appreciation to McREL staff members for their help in answering questions and solving special problems. Participants particularly liked the chance to interact with students, especially those students who had had previous experience with the IRA program.

Workshops by T_2's in Their Own School Districts

Three T_2's carried out their workshop plans in formal training prior to the 1972-73 school year. One T_2 was not able to carry out her planned workshop on a formal basis, but continued to provide discussion and informal training for one teacher. The participants were selected as described elsewhere in this report and attended the workshop under the following remuneration agreements.

1. T_2's to receive $250.00 for workshop period.
2. T_3's to receive $200.00 for workshop period.
3. These provisions did not apply to Installation A which received no financial help from McREL.

A description of each workshop site follows.

Site A - No formal workshop was conducted even though a workshop was planned for at least two teachers. A follow-up report by T_2 indicated a loss of interest among T_3's who had previously been interested. No special time had apparently been provided for the workshop and no financial remuneration given to the teachers. The T_2 did continue to provide training to one T_3 in the school with discussions after school. Several hours were spent in these discussions and, according to notes returned to T_1's, most of the ideas in the original workshop plan were covered during this informal training. No information was available on the training success of the T_3; the T_3 dropped the program in November 1972. During a follow-up contact with this T_3, the ideas in the program were praised as were the T_2's efforts to carry out the implementation of the program. Pressure from college bound students (mainly juniors and seniors) who were very "content-oriented" seemed to be the most important reason for discontinuing the program. However, administrative support was positive and was an important factor in the T_2's ability to carry out the program in her own classes. The T_2 was also able to coordinate a large amount of data from non-IRA teachers at this installation site, an important contribution to field test results.

Site B - An eight-day workshop was conducted by a McREL trained T_2 for four T_3's. The workshop, which required special planning due to a flexible modular schedule used in the school, was completed by all four T_3's. According to the T_2 workshop summary report, all participants worked together as a team until each T_3 understood the activities. They were able to meet weekly throughout the year because the school district allowed specific planning periods for such purposes.

A McREL staff member visited the installation workshop and reported that although progress seemed very satisfactory, some significant problems had occurred that were beyond the control of the T_2, including: divided responsibilities among the T_3's such as football practice; lack of video tape equipment; and inability to conduct micro-teaching. The inability to conduct micro-teaching was offset by the team teaching which offered an opportunity for teachers to observe and critique each other's teaching. A very high morale among the participants was reported, as was ability to deal with technical problems, enthusiasm for the program, and willingness to put forth strong effort to implement the program. The report also indicated a strong commitment from local school staff to support the program. Because of the team teaching situation, these five teachers were considered one for purposes of data analysis.

Site C - A 10-day workshop was conducted by a McREL trained T_2 for 3 T_3's Two teachers completed the workshop; the third teacher dropped out on the fourth day. All T_3's and the T_2 taught biology in the same building.

According to this T_2's report, submitted at the end of the workshop, the planned workshop agenda was followed except for the introduction of IRA program, which was replanned during the workshop session (a description was sent to McREL in "flow charts" describing the sequence and relationship of activities and was acceptable to McREL personnel). Also, micro-teaching was omitted partly because the T_2 felt that the T_3's were not "comfortable enough" with the activities to try them out in micro-teaching sessions. As a result, the micro-teaching was replanned for the eighth day, but was rejected in order to complete the 10-day workshop on time. During the McREL workshop, this T_2 seemed favorable to micro-teaching and the use of students to try out the activities. While it is very difficult to determine the effect of this omission of micro-teaching, the McREL staff felt it could be rather serious.

The T_2 reported that one T_3 met the criteria for all activities covered in the workshop and expressed doubt about the progress of the second T_3 and appeared hesitant to provide him with remedial help. This second T_3 dropped the program a short time after school began, giving the same reason a third T_3 gave when he dropped the program during the fourth day of the workshop: "I'm a traditional teacher and it required too much to change."

In addition to this T_2's general assessment of the workshop activities, he reported difficulties in the training similar to the ones the McREL staff had encountered during their Kansas City workshop. Besides problems with the workshop plans, he expressed specific concern for biology content, errors in the Teacher's Manual, loss of confidence in "educational methodologies," need for better description of grading system, and concern about using hemocytometers when many schools do not have them available. The errors, grading system description and use of complex equipment will be revised in the Teacher's Manual instructions. The T_2 suggested small numbers of participants such as he had in his workshop restricted group discussion. He suggested the Teachers' Manual be simplified with separate training sections. Finally, he concurred that a 7-day training session was feasible.

One McREL visit to this workshop was for the main purpose of monitoring the workshop progress and coordinating plans for data collection throughout the school year. The report following this visit did not obviate and seemed supportive of comments made in the T_2's report. Administrative support for the program also seemed very positive.

Site D - A 10-day workshop was conducted by a McREL trained T_2 for two T_3's who completed the workshop and succeeded in a follow-up implementation in their schools. One T_3 and the T_2 taught in one school and the other T_3 taught in a separate school.

The proposed plan of the workshop was very similar to the workshop conducted at Site C. Since the two groups were in the same district, they planned to conduct a joint workshop, but this plan was preempted by McREL Evaluation Design Committee who decided it would contaminate the evaluation program to have a mixture of two T_2's influence in the same workshop. Despite this they conducted joint sessions on two successive days. The workshop included micro-teaching. The participants reacted positively and few suggestions were made for improvement, except for these concerns over manual interpretations, difficulty in arranging and using audio visual equipment, and receiving materials on time. Manual interpretations and audio tape problems have been generally corrected; other problems have been subsequently resolved within the school district.

One visit to the workshop showed no serious problems except technical difficulties such as how to grade a team when one member is missing. It is strongly suspected, although not definitely stated, that workshop discussion would have been enhanced with a larger group of participants. Administrative support for this installation seemed very positive.

Continued "T_2-T_3" Relationship-Throughout the School Year

Site A - Informal discussions were conducted by T_2 with T_3 up to the point of withdrawal from the program. The T_3 subsequently reported to McREL staff that his relationship with the T_2 was cordial and that he had dropped the program for other reasons.

Site B - The five teachers at this site used a team teaching approach coupled with a modular schedule designed to maximize a student's exposure to a variety of teachers and other students. Because of the team teaching and limited in-class time, teachers closely followed a master schedule of activities and lessons. Part of the teacher's schedule was a weekly two-hour planning period. In addition, periodical meetings were scheduled after the regular school day. T_2 reports and T_1 on-site monitoring reports indicate that these teachers worked very well together and conferred with one another daily.

Site C - Follow-up contacts by the T_2 with T_3's were carried out on a formal and informal basis. The T_2 assumed a title as Coordinator of Inquiry Role Approach Training at his school. According to reports made on several follow-up meetings conducted during the year, there appeared to be a much closer personal relationship between the T_2 and one T_3 who continued in the program than between the T_2 and the T_3 who dropped the program. This relationship persisted throughout the school year.

Site D - Follow-up contacts by the T_2 with T_3's were almost entirely on an informal basis. The T_2 and one T_3 worked together in the same department and were able to discuss and plan the program extensively. The other T_3, located in another school, maintained a close relationship with the T_2 by telephone conversations. According to reports, the T_2 assumed other non-IRA responsibilities during the second semester; however, by this time the T_3's were apparently implementing the program in a satisfactory manner.

Site E - T_3 worked alone. McREL personnel maintained contact on a minimal basis by mail or telephone. The T_3 was encouraged to call McREL personnel in case of emergency or for clarification of Teacher's Manual instructions or testing; otherwise, the T_3 was very much on his own. One visit was made early in the school year, mainly for the purpose of collecting data and coordinating testing.

Site F - T_3 worked alone. McREL personnel maintained a minimum of contact by mail or telephone on procedural matters. Two classroom visits were made early in the school year, mainly to clear up audio visual and data problems.

Site G - T_3 worked alone. Since this teacher had had some experience with IRA methods and activities three years ago, but no workshop training on the latest materials, three visits were made to the classroom to brief the T_3 on how to use the Teacher's Manual. A total time of less than two hours for the whole year was spent briefing this T_3 plus occasional telephone conversations with her to clarify manual instructions.

Site H - T_3 worked alone. This T_3 was a co-developer of the program so classroom visits were made mainly to discuss field test problems and help coordinate testing programs in the school district.

Communication by T_1's With all Participants Throughout the School Year

Feedback forms, called Teacher's Logs (See Appendix), and onsite visits added to a continuing communication process among McREL staff (T_1's) and field test participants. All T_2's and T_3's were asked to fill out Teacher's Logs after completing each activity. McREL staff would examine them and respond if necessary. The logs were divided among T_1's who would communicate periodically with teachers on progress being made in implementing the program. The T_1's paid special attention to such things as amount of time spent on activity, whether the activity was completed, what percent of students reached the objectives, and kind of modifications made in presenting the activity. Teachers were encouraged to suggest how the activity might be revised to increase its effectiveness. T_1's reacted to the logs by answering specific questions raised by the teacher or giving information about how other teachers had carried out the activity successfully.

While the Teacher's Logs gave the teachers' perceptions of the activities, a pair of instruments was a source of data on students' perceptions of the activity, resulting in valuable information on whether selected teaching practices had been carried out during the implementation process. At the end of Theme I, (January, at most sites) the Views and Preferences-Form C instrument and Classroom Activities Questionnaire (CAQ) (see Measurement Instruments section) were given to students. This data was analyzed and results were reported back to the teachers. The teacher could then compare his perceptions of what had happened in Theme I with his students' perceptions. For the T_1's this information provided valuable input to supplement the results given in the Teacher's Logs. This dual method of determining the degree of implementation was an important aspect of the field test.

Other communications with T_2's and T_3's were visits to the classrooms. Those conducted in schools near Kansas City were reported in the last section; the others are described here.

Site A - No visits were made due to lack of funds for long distance travel.

Site B - Visit made in April 1973. The major objectives of this visit were: to identify problems of implementation; assess teacher and student attitudes towards IRA; discuss plans for the future of IRA.

Developing continuity among students and teams while operating within a flexible modular scheduling was found to be a major problem because it was very difficult for teachers and students to coordinate individual work, small group discussions, and large group discussions. Also, some students showed concern over too much paper work, not covering enough biology, and some dislike for the roles; but indicated respect for the goals and methods of IRA and thought IRA was better than the former method of learning biology, as described by their friends.

The teachers were pleased with IRA because they were able to compare IRA with traditional methods previously used and could compare results over two or more years. Increased discussion was a favored point for the IRA program.

Plans were to be made to use IRA in all senior high and mid-high biology classes. University of Oklahoma staff, after frequent visits to IRA classes, redesigned a methods course (soon to be offered) which will incorporate IRA methods.

Site C - Visit made in March 1973. A one-day visit was shared between a T_2 and a T_3 at this site. Both teachers were carrying out the implementation processes satisfactorily. Procedures for holding discussions and special orientation section for new teachers were suggested improvements they gave for the Teacher's Manual. Students reacted positively to IRA. One innovation worked well for this team--the T_2 stayed a week ahead of the T_3 in scheduling his activities, so that the T_3 would benefit from his experience of what went well, what didn't work out, and what to do about it. The T_2 was also actively involved in workshops with other than biology teachers in an effort to adapt some IRA ideas to teacher training and curriculum development in other subjects and departments. He was also somewhat successful in interesting local colleges to use IRA ideas in their methods classes.

Site D - Visit made in March 1973. A one-day visit was shared among one T_2 and two T_3's at this site. As was done at Site C, the T_2 and a T_3 stayed ahead of the other T_3 for scheduling their activities. The model of this new T_2-T_3 relationship at Sites C and D was reported quite effective. Some difficulty was experienced with certain inquiry guides, a need for more materials on roles, and using the video tapes on roles. The students in this installation generally liked and were receptive to the IRA roles except in the one class where, apparently, some students desired a content-oriented, memory type of course. The T_2 at this site cooperated with the T_2 at Site C in a joint presentation of IRA to local educational groups.

Site E - Visit made in January 1973. The T_3 at this site was very conscientious about developing students' positive attitudes towards themselves and the program. The students seemed to work well in the activities. Some concern was reported about the slow pace of the program and that continuity of the program would be lost if too much time was taken on certain activities. The T_3 seemed receptive to helping to install the program in other schools and departments and was instrumental in hiring another teacher, who had had previous IRA experience. The administration at this site was especially supportive.

One major reason for this visit was to collect data on "non-IRA" teachers. The T_3 coordinated the data collection process and returned this data to McREL. These efforts were comparable to those carried out by T_2's in Installation Sites A, C and D.

Site F - Two visits made--one in late fall; one in winter of school year 1972-73. This visit was made mainly to coordinate video tape schedules and equipment, and testing; however, some time was spent on visiting the classroom. The easy going manner of T_3 in working with students was seen as favorable, but some concern was expressed as to whether the pace, or scheduling, of activities was too slow, and whether the students really understood the IRA ideas and concepts. Administration seemed neutral to the existence of the IRA program.

Site G - Three visits made--December, 1972, February and April 1973. Classroom visits were made because this T_3 had not completed the McREL workshop, even though she had had previous experience using the first IRA materials in 1968-69. The T_3 reported little difficulty in following the Teacher's Manual but when problems did occur (principally on the testing procedures) she would call McREL. The students seemed receptive to the IRA program. Administration was generally in support of IRA.

Site H - Visits made 15 to 20 times during the school year. The main reason for these visits was to determine how these students reacted when taught by this teacher who had previous experience with IRA. This experienced teacher suggested ways inexperienced teachers could better overcome their immediate problems and made revisions in the Teacher's Manual. One major difference between this teacher and the others was his ability to adjust the scheduling of the activities to the students' abilities. He carried out more remedial work during this year than in the previous years. The students were positive toward the program and completed extra class activities such as science fair projects.

Experienced IRA Teachers - Several visits were made to classrooms of three teachers with previous experience with the program. These teachers have made various adaptations of IRA to accommodate the use of other textbooks, a second course in biology, etc., but have retained the basic IRA concepts. Suggestions by these teachers have been valuable in rewriting program materials. These teachers continued to give their time and services to the development of IRA even though no funds were available for their compensation. This group of teachers has been previously described in the Description of Participants section.

REFERENCES

1. "Development of Views and Preferences - C," Unpublished paper presented at NARST, 1973, Lowel A. Seymour and Richard M. Bingman.
2. Steele, J., House, E., and Kerins, T. "An Instrument for Assessing Instructional Climate Through Low Inference Student Judgments," AERA Research Journal, May, 1971.
3. Biological Sciences Curriculum Study, Comprehensive Final Examination, Form J and K, New York: The Psychological Corporation, 1965.
4. Koos, E. M., (Principal author)"Explorations in Biology," Topics 1 through 8, Mid-continent Regional Educational Laboratory, 1969-72.
5. Steiner, H. Edwin, "A Study of the Relationship Between Teacher Practices and Student Performance of Selected Inquiry Process Behaviors in the Affective Domain in High School Biology Classes," Unpublished Doctoral Dissertation, University of Texas, Austin, 1970.
6. Bennett, G. K., Seashore, H. G., and Wesman, A. G., Differential Aptitude Tests, New York: The Psychological Corporation, 1959.
7. Krejcie, Robert V., and Morgan, Daryle W. "Determining Sample Size for Research Activities," Educational and Psychological Measurement, Vol. 30, 1970, p. 607.
8. "Small-Sample Techniques," The NEA Research Bulletin, Vol. 38, 1960, p. 99.
9. Manual for the Differential Aptitude Tests, The Psychological Corporation, New York, N.Y. [illegible]0017.

MEASURING INSTRUMENTS

Teacher's Log*

This instrument was designed to help determine the implementation of the Inquiry Role Approach program. The instrument was written by McREL staff to supply information specific to each activity not previously secured. The data was to be recorded on the log at the end of each activity by each teacher participant. This instrument and procedure were described in the training workshops.

The data recorded by the teachers on the logs included: activity number, whether the activity was completed and to what extent, time spent, modifications and explanations in activity procedures, reactions to the sequence, the pre- and in-class instructions, and the student materials, an estimation of the percentage of students meeting criteria for objectives, and, information on how each activity could be improved. Data from this source were verified in some instances by other types of communications, e.g., telephone conversations, direct observation, interviews, and letters. Based on these other sources of information it was determined that the reports on the Teacher's Log were reliable. The items elicited information desired and were thus valid; but, of course, extraneous data were supplied and in some instances participants chose not to use the log as their means of communication.

Views and Preferences - C*

The V&P - Form C contains 50 items which were selected from 143 items of Views and Preferences - Forms A & B. The items were mainly selected on the basis of whether or not they discriminated between 700 IRA and 520 non-IRA students. The non-IRA students in this sample were enrolled in BSCS biology classes and used a standard textbook laboratory approach.

The data for the two groups were analyzed by calculating a chi-square for each item. The items selected, the level of significance, and the percent who chose the desired response are recorded in the Seymour and Bingman paper.[1] Differences between the two groups were significant for 49 items and another item was retained because IRA students met the criterion level and it was deemed to measure an important aspect of the Inquiry Role Approach program.

In order for an item to be retained in Form C, IRA students had to score in the desired manner and to the pre-determined criterion level. The professional judgment of the staff determined these, and they are recorded in the aforementioned paper.

There are three major dimensions of Form C: social, cognitive, and class procedures. The items pertaining to these respective dimensions are 1-16, 17-36, and 37-50.

---

* See Appendix for sample.

A mean score for each item of sets of items can range from 1.0 to 5.0 where 1.0 indicates an undesired IRA response and 5.0 indicates a desired response. A mean score of greater than 3.5 indicates that a majority chose the desired response.

The test-retest reliability was computed for Views and Preferences and was found to be 0.80.[2]

The items for Forms A and B were written by McREL staff and consultants, and were judged to measure various implementation aspects of the IRA program. A further description of this process is provided in the Seymour and Bingman paper.

Part of the reason for using students via this instrument to help judge implementation was the rationale provided by Steele.[3]

> "It was judged that the most accurate estimate of cognitive emphasis and positive learning environment could be obtained using sensitive and perceptive observers who would be in the class frequently and who were trained in using systematic procedures to collect the data. This procedure is too costly. The training, time, and support demands prohibit its use . . . However, two sources of untrained observers exist in any classroom: the teacher and the students."

Teachers and students were equally used to determine the extent of IRA implementation. The four-part determination contained two implementation aspects reported by teachers via the Teacher's Logs and two aspects reported by students--one was the Class Activities Questionnaire (CAQ) and the other was the Views and Preferences - C.

V&P-C was used to measure students' perceptions of the class procedures, to social and cognitive aspects of IRA classes.

Class Activities Questionnaire (CAQ)*

The CAQ[4] is a 25-item instrument administered to both students and teacher. It asks students to agree or disagree on a 4-point scale to statements describing general kinds of activities which characterize their class. These activities imply either levels of thinking or affective classroom conditions. Each item is paired with another item to compose a factor; sixteen factors yield a revealing profile of the class. In addition, subscores are derived by clustering factors into the four dimensions of Lower Thought Processes, Higher Thought Processes, Classroom Focus, and Classroom Climate. The cognitive dimensions of Lower and Higher Thought Processes represent a dichotomy strongly supported in validation studies of Bloom's Taxonomy. The Classroom Focus dimension assesses whether the focus is on the teacher as information-giver with students having a passive role, or on the students having an active role in the class. The Classroom Climate dimension assesses attitudes and feelings, such as how relaxed and open the class is and the amount of involvement of students in class activities.

---

* See Appendix for sample.

Extensive field testing was used in developing the 25 items. Interviews with students were conducted and revisions were made to ascertain if the words used were understood and if statements were appropriately interpreted. Grade six was determined to be the lowest grade level at which students could understand the items and make the judgments called for. In addition to field testing with children, the cognitive items were classified by judges familiar with Bloom's taxonomy to determine whether the items were seen as appropriate for the intended taxonomic categories.

The teacher would be a poor source from which to obtain information about the actual emphases occurring in the classroom. However, the teacher is the most direct source from which to obtain data on what is intended to be emphasized. It is for this purpose that the teacher is asked to respond to the CAQ. The teacher reports his intended emphasis and also predicts what the students as a group will say. The teacher can then compare these responses with the actual emphasis perceived by students. Students are in a much better position to report on the emphasis actually given to various class activities. Moreover, the nature of the instructional climate depends in part on the way it is perceived by the students themselves. Not every student is an accurate observer, however, it is the consensus of student judgments that is of concern. A system of consensus scoring, rather than simple mean or median scores, is used.

The Horst formula was used by Steele for estimating reliability from the within class and between class variances was used. The reliability estimates for each of the four major dimensions as well as for each of the 16 individual factors were obtained. Fourteen of the 20 correlations are above .80 with only one falling below .65. A pilot study has been conducted to explore the stability of response over time. The test-retest reliability coefficients for each of the four dimensions are .67, .91, .59, and .89, respectively.

The CAQ instrument was developed and used on the basis of a logical design. To ascertain the degree to which the data supported this structure, a principal component analysis of items 1 through 25 was conducted. The statistical components provide substantial support for the logical construction of the instrument.

Side two of the CAQ has three questions:

> 28 - "List the three best things about this class from your point of view."
>
> 29 - "If you could change three things about the class, what would they be?"
>
> 30 - COMMENTS: If you have any comments, please write them below.

Students responded to these questions at the end of Theme I and at the end of the school year. These responses were categorized and are reported in Data Analysis section of this report.

Comprehensive Final Examination (CFE)*

The CFE[5] is, as its title states, designed as a comprehensive examination of the achievement in biology attained by students in a first-year secondary level biology class. Specifically the instrument has been designed for the BSCS courses using any of the three BSCS textbooks; however, it is seen as applicable to other modern biology curricula as well. Two equivalent forms, J and K, have been developed.

Validity: The validity of the CFE has been primarily determined by the judgment of subject matter specialists and the supervisors of the writing teams for the three BSCS texts. In this manner the instrument has been judged to be valid in terms of covering the content of the three texts. In addition, validity was studied by determining the correlation between student scores on the CFE and on each of the four Quarterly Achievement Tests designed to accompany the three text versions. The coefficients of correlation range from .63 to .82.

Reliability: Both internal consistency of each form and correlation between forms have been studied. Using the Kuder-Richardson 20 procedure with a sample of 740 cases, coefficients of internal consistency ranging from .76 to .86 were found, with a median coefficient of .82 for Form J and a median coefficient of .84 for Form K. Coefficients of correlation for scores obtained on Form J and Form K have been found to range from .72 to .85 with a median coefficient of .79 (N = 2500).

Norms: In addition to validity and reliability studies, data from a national sample of 11,092 students, taking both forms of the test, have been used to establish norms or percentile rankings equivalent to each raw score possible. These are very useful to teachers in evaluating student performance. A raw score of 27 or 28 places a student at the 50th percentile in the norm population. It should be noted that the mean DAT-Verbal Reasoning + Numerical Ability composite score for the CFE norm population was at the 70th percentile.

Biology Student Behavior Inventory (BSBI)*

The BSBI[6] is a 39-item instrument assigned to measure the frequency of occurence of specific student behaviors indicative of four attitudes considered necessary for cognitive inquiry -- curiosity, openness, satisfaction and responsibility. The student is presented with a situation and a selection of possible behaviors or actions that could be taken in that situation. The student indicates what he would do in this situation by selecting one behavior. The preferred responses (which receive a score) are behaviors indicative of one of the four attitudes given above. Four sub-scores or sub-scales are therefore determined. 11 of the items are used to determine the curiosity subscore; 17 are used for the openness subscore; 7 for the satisfaction subscore; and 4 for the responsibility subscore.

---

* See Appendix for sample.

Validity: Validity has been studied in three ways--by a panel of nine judges; by correlation of student item scores with student subscore scores (this was used primarily to confirm categorization when judges did not show a high percentage of agreement); and by correlations with a second instrument (Observational Record of Affective Behaviors, ORAB) which measured the same attitudes utilizing fewer behaviors and an observational approach.

The judges' agreement has been reported as the percent agreeing with the test author. In keying the BSBI items to one of the four attitudes, 67 percent agreement or higher was found for 33 of the 39 items; average percent agreement for all 39 items was 83 percent.

To confirm the judges' findings and in particular to evaluate the categorization of the six items which showed low (below 67 percent) percentages of agreement, a Pearson product-moment correlation coefficient was determined for each item. Student item scores were correlated with each of the four student subscale scores. This process confirmed the validity of the previous categorization of items.

Finally Pearson coefficients were found for student scores on three sub-scales of the BSBI (curiosity, openness and responsibility) and total BSBI score (using only three sub-scales) correlated with the same three subscores on the ORAB and the total ORAB score. The curiosity subscales had a correlation coefficient of -.45; the openness subscales, .88; the responsibility subscales, .75; the total scores, .83 (for significance at the .05 level, $r \geq .75$). The low curiosity subscale correlation appeared to be due to the fact that the ORAB measured primarily only one behavior indicating curiosity while the BSBI measured five behaviors; thus the two instruments were not measuring the same behaviors and low correlation could be expected. It should also be noted that the ORAB contained a non-inquiry subscale; this subscale showed negative correlation with each BSBI subscale and the BSBI total score.

Reliability: An estimate of the reliability of each subscale was determined using a split-half technique. Pearson product-moment coefficients of correlation were found and adjusted using the Spearman-Brown formula (Guilford, J.P., Fundamental Statistics in Psychology and Education. N.Y.: McGraw-Hill Book Co., 1965. p. 457).

With a student N of 1153, the following values were computed: curiosity, $r = .67$; openness, $r = .68$; satisfaction, $r = .71$; responsibility, $r = .37$; for significance at the .01 level, $r \geq .07$. With a class N of 48, the following values were computed: curiosity, $r = .78$; openness, $r = .68$; satisfaction, $r = .86$; responsibility, $r = .51$; for significance at the .01 level, $r \geq .37$. A Cronbach alpha was also computed to determine internal consistency for each subscale. The alpha values: curiosity, $\alpha = .65$; openness, $\alpha = .71$; satisfaction, $\alpha = .66$; responsibility, $\alpha = .43$ ($N = 1153$). BSBI A, B, C, D, and total had Cronbach alpha values of 0.55, 0.78, 0.68, 0.37, and 0.84 respectively. The sample was the experimental group.

BSBI was used to measure curiosity, openness, satisfaction, and responsibility.

Explorations in Biology (EIB) 1 and 2*

The EIB series is a set of eight simulated problem-solving instruments designed to measure cognitive inquiry skills. These instruments have been developed in the period of 1969-72 as a component of the Development of Inquiry Skills Program of McREL. The instruments are designed to measure the following inquiry skills:

14 Inquiry Objectives -

1. Identifying a phenomenon to investigate.
2. Identifying the question arising from the identification of this phenomenon.

3a. From a list of readings, selecting and evaluating reports possibly yielding useful information about the event noted. (Explorations 2, 3, 4, 5, and 6)**

3b. From relevant readings on the problem presented, decide if given hypotheses are tenable. (Explorations 1, 7, and 8)**

4. Differentiating likely causes of this event from unlikely causes.
5. Selecting a single hypothesis to investigate.
6. Selecting an array of methods appropriate to the investigation.
7. Identifying the independent variable to be studied.
8. Identifying conditions required for conducting a laboratory study on this topic.
9. Choosing a plan which would yield data affording a test of the hypothesis.
10. Identifying assumptions necessary for interpretation of data resulting from carrying out the plan.
11. Identifying the data which would result from carrying out this plan.
12. Identifying justifiable conclusions from data associated with a class experiment on this topic.
13. From a heterogeneous list of questions, identifying new questions which might arise as a result of carrying out this investigation.
14. Integrating results of this study with those reported by other investigators in related areas.

---

* See Appendix for sample.

** EIB's 1 through 6 were developed to measure the above set of Objectives including 3a but not 3b. The format for EIB's 7 and 8 was slightly changed from an earlier format used for EIB's 1 thru 6. With this new format, Objective 3b was substituted for 3a. In 1972 EIB 1 was revised into the new format; minor changes in EIB 2 resulted in use of both Objectives 3a and 3b for this instrument.

In the 1972-73 field test, EIB 1 was used as the pre and posttesting instrument for assessing cognitive inquiry skill student outcomes and pre-to-post gain. EIB 2 was used to assess interim inquiry skill development, given as part of Theme I Assessment in Activity 121.

Validity - Objectives were selected for the EIB's based on studies by Burmester,[9] Kaplan,[10] Suchman,[11] and Taba.[12] With the completion of the detailed McREL-BSCS set of inquiry objectives,[13] studies were made to learn the extent to which EIB items would be referenced to similar objectives listen in the Inquiry Objectives in the Teaching of Biology document. These studies were previously reported by Koos.[14]

Changes in the Explorations in Biology since these studies were undertaken have been primarily format changes and changes in wording to clarify directions and meaning. However, the inquiry objective content validity was reviewed in the summer of 1972 by two McREL staff members and a teacher-consultant.

Working independently, each judge keyed the test items using: the 14 EIB objectives, a category for items in which sequence of test steps were chosen, and a category for items not related to any of the objectives or step choice. Disagreements were found for less than 15 percent of the items. In most cases, disagreements resulted from misreading, or misinterpreting, the test items or directions. In all cases, disagreements were discussed and concensus reached for keying the item. In addition, the EIB objectives were categorized by the judges as related to six major areas of cognitive inquiry behaviors. The following table presents the categorization of the objectives and the items in EIB 1 and 2 keyed to each objective:

EIB 1 & EIB 2 ITEMS KEYED TO INQUIRY OBJECTIVES		ITEMS ON EIB 1-A	ITEMS ON EIB 1-B	ITEMS ON EIB 2-A	ITEMS ON EIB 2-B
AREA I - Formulating a Problem	Objective 1	1		80	
	Objective 2	50		50	
AREA II - Searching for Information	Objective 3a			6-15	
AREA III - Formulating Hypotheses	Objective 3b	8-17		26-35	
	Objective 4	51-60		51-70	
	Objective 5	48,49		48,49	
AREA IV - Designing an Experimental Study	Objective 6	36-45		36-45	
	Objective 7		1		1
	Objective 8		2-6		2-6
	Objective 9	46,47	17,43-47	46,47	17,43-47
AREA V - Interpreting the Data or Findings	Objective 10		7-16		7-16
	Objective 11		18-42* 48-57		18-42* 48-57
	Objective 12		58-67		58-67
AREA VI - Applying and Synthesizing Knowledge	Objective 13		68-77		68-78
	Objective 14		78		79
Step choice items	(not scored)	2-7		1-5	

* Students choose one set of 5 items to respond to in the 18-42 group.

Based on this categorization and assignment of test items to objectives, sub-scores for each of the six inquiry areas can also be determined in scoring the EIB's.

Construct validation studies have been made to compare EIB 1 with BSCS Comprehensive Final Examination,[15] Differential Aptitude Test-Verbal Reasoning & Numerical Ability and Abstract Reasoning,[16] California Basic Skills Test,[17] Iowa Tests of Basic Skills,[18] Scholastic High School Placement Test,[19] and Watson-Glaser Critical Thinking Appraisal.[20] A Pearson product-moment correlation of .63 was found between EIB 1 and the DAT-Verbal & Numerical scores. Minimal correlations have been found between EIB 1 and DAT-Abstract Reasoning and Watson-Glaser Critical Thinking.[21] Other correlations were found to be very low. "...construct validity is offered for those EIB 1 items which tap cognitive operations involving verbal formulation of biological problems, verbal interpretation of non-verbal data, and analysis of quantitative information presented in tabular or graphic form. This suggests that the intellectual factors of verbal reasoning and numerical ability are factors basic to successful inquiry."[22]

Reliability - The developmental 1969 version of EIB 1 was shown to have a coefficient of internal consistency (Kuder Richardson 20 procedure)[23] of .96 when tested with a heterogeneous group of 451 students; .74 when tested with a more homogeneous group of 150 students.[24]

The later 1970 versions of EIB 1 and 2 were tested on several occasions in the spring of 1970 and in the 1970-71 school year. Coefficients of internal consistency (Cronbach alpha)[25] ranging from .40 to .86 and averaging .73 were found for EIB 1; coefficients ranging from .75 to .99 and averaging .87 were found for EIB 2.[26]

While reliability was adequately demonstrated by these analyses, the EIB 1 format was revised during the summer of 1972.[27] Major changes involved the items keyed to Objectives 3a and 3b. Objective 3b was substituted for 3a, and related items were revised or replaced. This change was made to insure that all students were provided the same background information on the topic; formerly, readings from related science literature were optional. In addition, the items keyed to Objective 4 were reduced from 20 to 10. Some item numbering changes in Part A were also made. In order to establish the degree of reliability of the 1972 revised instrument, coefficients of internal consistency (Cronbach alpha) were determined for the total scores and part scores I, III, IV, V and VI of EIB 1 using the posttesting data from the IRA field test students. These coefficients are presented in the following table.

	COEFFICIENT OF INTERNAL CONSISTENCY	N
EIB-1, Part I	-.2385	1,005
EIB-1, Part III	0.6157	1,005
EIB-1, Part IV	0.8327	1,005
EIB-1, Part V	0.8549	1,005
EIB-1, Part VI	0.8807	1,005
EIB-1, Total score	0.8690	1,005

It should be noted that the scoring keys for EIB 1 and 2 were revised during 1972. Previous scoring keys had been devised by the test author and had not been reviewed by others. In discussing aspects of the 1972 revision of EIB 1, it was found that IRA staff members disagreed with the suggested scoring of some items. A more thorough review was planned with five McREL staff members acting as judges. New scoring keys, reflecting consensus among the five judges, were developed. The degree of difference between the original author's key and the revised key can be determined from the following table.

COMPARISON OF EIB-1 and EIB-2 SCORING ON ORIGINAL AND REVISED SCORING KEYS.

	TOTAL POSSIBLE RESPONSES	SCORED RESPONSES, ORIGINAL KEY	SCORED RESPONSES UNCHANGED	SCORED RESPONSES, SCORING DELETED	UNSCORED RESPONSES UNCHANGED	UNSCORED RESPONSES, SCORING ADDED	PERCENT AGREEMENT $\% = \frac{3+5}{1} \times 100$
EIB-1A	158	51	34	17	94	13	$\frac{34+94}{158} \times 100 = 81.0\%$
EIB-1B*	220	82	76	6	128	10	$\frac{76+128}{220} \times 100 = 92.7\%$
EIB-2A*	175	66	42	24	89	14	$\frac{42+89}{175} \times 100 = 74.9\%$
EIB-2B*	222	65	58	7	135	22	$\frac{58+135}{222} \times 100 = 86.9\%$
TOTAL	775	264	210	44	446	59	$\frac{210+446}{775} \times 100 = 84.6\%$

* All optional items included.

The author's key had used a "weighted" scoring system. Scored responses could be awarded either 2 or 1 point. Criteria for weighting the value of responses appeared to include difficulty of the item, degree of accuracy of response (when more than one response to an item was scored), and whether the response was negative rather than positive (the author felt a negative response to an item was psychologically more difficult to make).

The panel of judges felt that these criteria were not consistently applied. The author has not specified a systematic approach for assigning weighted scores. The judges, therefore, decided to delete weighting of scores as much as possible--weighted scores are used in the revised key only for optional sections when necessary to maintain equal chance scores for each option presented.

All EIB 1 data in this report utilizes the revised key. EIB 2 was scored by teachers at the end of Theme I. Theme I materials included the original scoring key for teachers' use. Some teachers had completed Theme I before this error had been noted and revised scoring keys were supplied.

Maximum and chance scores for EIB 1 and 2, and part scores for each are given in the following table.

	ORIGINAL AUTHOR KEY				REVISED PANEL KEY			
	EIB-1		EIB-2		EIB-1		EIB-2	
	MAXIMUM SCORE	CHANCE SCORE	MAXIMUM SCORE	CHANCE SCORE	MAXIMUM SCORE	CHANCE SCORE	MAXIMUM SCORE	CHANCE SCORE
Part I	4	1.20	2	1.60	2	.40	1	.20
Part II	0	.00	7	3.60	0	.00	4	2.00
Part III	41	14.08	46	21.42	22	7.07	26	11.36
Part IV	45	17.15	42	13.80	24	11.30	24	10.97
Part V	60	27.00	42	18.00	40	17.50	35	16.25
Part VI	21	10.10	18	8.60	10	4.70	12	5.90
TOTAL	171	69.53	157	67.02	98	40.97	102	46.68

Social Skills Checklist*

This instrument is designed to measure student performance of social skills in the areas of communication, coordination, and role performance. The instrument is completed individually and by teams, and this composite rating becomes the assessment of social skills for each student. About 55 minutes are required to administer this 55-point checklist. A copy is found in the appendix of this report. The checklist was developed by McREL staff with assistance from two experienced IRA teachers who had given it prior testing. The experienced IRA teachers confirmed that these were the salient elements of the domain termed social skills. To determine reliability of the checklist further study should be considered.

Attitude Checklist*

Chapter 5 of Inquiry Objectives in the Teaching of Biology (IOTB)[28] provided the theoretical foundation for this checklist. Scores of objectives in the area of affective or attitudinal qualities of inquiry behaviors are delineated. The attitudes selected for inclusion were:

- A willingness to participate in inquiry activities.
- A willingness to assume responsibility in inquiry activities.
- A willingness to cooperate with other student to complete an inquiry activity.
- A willingness to change ideas and evidence when it is necessary.
- A willingness to admit his mistakes.
- A willingness to look for additional data and evidence.
- A concern for issues in the public domain.

Team and individual completion of this 65-point instrument takes approximately 50 minutes. The 13-item instrument is found in the appendix of this report. The items and objectives were chosen by five McREL members and two experienced IRA teachers to be the ones emphasized in Theme I, and the more predominant ones in these early (first semester) activities. Further study of the reliability of each item is in order. Most of these items have been previously used in other forms, and the data supports the consistent response on the part of students.

---

* See Appendix for sample.

Biology Content Test*

This 72-item, multiple choice test was developed from Resource Book of Test Items for Biological Science - An Inquiry Into Life, 2nd edition, BSCS, Educational Programs Improvement Corp., 1971. The items selected are given in Section 121-2 of Theme I. Items were selected for text chapters 1-7, 12, 18, and 36-38. Information and definition type items were 34.7 percent of the total while 65.3 percent of the items were categorized as application and inquiry processes items. Validity and reliability were not reported by the developers.

The items selected were judged by three program development specialists, one research and evaluation specialist, and two experienced IRA teachers to measure validly the biology content of Theme I.

Understanding Role Responsibilities*

This 20-item, 4-choice multiple choice instrument (Section 121-3 of Theme I) was developed over a three year period to measure students' ability to properly identify which role was responsible for a particular statement given. Revisions have occurred over the years to keep the items consistent with the development of instructional materials on roles. Development has also continued because of the responses by teachers to the wording of various statements.

Validity was established by the professional judgment of McREL staff. These judgments were supported by discussing the meaning of items with teachers participating in IRA development and their students. Experienced IRA teachers note that students score higher on the instrument after repeated instruction on role responsibilities.

Even though validity and reliability can not be substantiated with complete evidence, a report of the development of this instrument is in order.

In the 1970-71 IRA materials the instrument used to measure role understanding was Differentiation Between Roles-35c, a twenty item instrument. The validity of the items was judged by interviewing and observing teachers and students (reported in "Summary of Evaluation Meeting," January 20, 1971, John Anderson, Richard Bingman, and Charles Dowler). This instrument was revised and designated Form 35d; Form 35d included 13 items identical to items in 35c and seven modified items (described in Report on AY '70-'71 Evaluation and Revision of IRA Component of DIS). A test-retest reliability study of the 13 identical items from Form 35c and Form 35d indicated a rank correlation coefficient (Spearman) significantly different from zero at the 0.01 level. The items were very stable in this test-retest situation. The percent correct deviations ranged from 10.6 to 1.5. Ten of 13 scores deviated no more than 5.1 percent. These statements are based on an N of 6 teachers and 580 students (also described in the '70-'71 evaluation report).

During 1971-72 the instrument was changed from a 20-item to a 33-item instrument and was termed Assessment of Role Functions-Form 21. (Form 21 was used with IRA materials keyed to the BSCS Blue Version textbook; the same instrument was designated Form 26 in the materials keyed to the BSCS Green Version textbook.) The longer instrument was found to be less useful due to its length and an unequal distribution of items related to each of the four roles used in the small groups. Further, it was found that some of the items were ambiguous and could be attributed to more than one role. In revising this instrument for the 1972-73 field test, the professional judgment of three program development specialists, one research and evaluation specialist, and two experienced IRA teachers was combined to select the 20 items (five for each role) which were most obviously characteristic and exclusive to a role. The revised form was designated Understanding Role Responsibilities 121-3.

---

* See Appendix for sample.

Form 35d is very similar to Section 121-3. Seventeen of 20 items (including the 13 studied for reliability in 1971) are identical except for the absence of quotation marks on Section 121-3. The difference between the two forms is in the wording of three items, the description preceeding the items, and the manner in which answers are indicated. The last two differences are very minor as one could note by reading the two forms. The three items that are different are shown below. The minor changes were made because of changes in the instructional materials.

Form 35d-Item 6:
"Do the rest of you agree with the summary of the Main Idea?"

Section 121-3-Item 6:
"Do the rest of you agree with the summary of the generalization?"

Form 35d-Item 12:
"The main idea in this statement deals with the way group members respect each other."

Section 121-3-Item 12:
Team is substituted for the word group.

Form 35d-Item 10:
"How is our Main Idea related to the Search for Overall Idea?"

Section 121-3-Item 10:
"How is our generalization related to the problem?"

Differential Aptitude Test (DAT)*

The DAT[29] is a battery of instruments designed to measure student aptitude in eight areas. Two of the eight instruments -- Verbal Reasoning and Numerical Ability -- are often used together as a measure of general learning ability (DAT manual, p. 1-7). Only these two instruments of the DAT battery have been used in the field test. These were administered in the fall of the year to establish a base for comparison made between groups in the field test and for comparing the field test group as a total with outside populations.

---

* See Appendix for sample.

Validity: A large number of studies have been preformed relating course grades for various subjects to DAT scores. It is adequate for our purposes to note that of the coefficients of correlation computed for science grades compared to the nine DAT scores (8 instruments and the Verbal Reasoning + Numerical Reasoning composite score), the highest coefficients were found for Verbal Reasoning (.45), Numerical Ability (.44) and the VR+NA composite (.52).

Validation by a 3-1/2 year longitudinal study has also been performed. This study indicated that DAT scores remain predictive of student performance over a long range. For example, DAT VR and NA scores from students 8th grade (mid-year) correlated well with general science grades achieved at end of 8th grade (VR - science grades, r = .64; NA - science grades, r = .59); these 8th grade DAT scores still correlated well with science (physics) grades achieved at end of 11th grade (VR - physics grades, r = .59; NA - physics grades, r = .60).

A most important means of validating the DAT is in appraising its predictive ability of student results on achievement tests. Some examples of the coefficients of correlation found between DAT-VR, DAT-NA and DAT-VR+NA scores and various achievement tests are given in the following table:

TEST	COEFFICIENTS OF CORRELATION							
	BOYS				GIRLS			
	N	VR	NA	VR+NA	N	VR	NA	VR+NA
Iowa Test of Basic Skills - Form 1 - Reading Comprehension	125	.62	.61	.69	117	.68	.61	.73
Arithmetic Total	125	.71	.69	.80	117	.63	.75	.76
Iowa Tests of Educational Development - Form Y4-FL Composite	93	.91	.85	.92	79	.89	.76	.89
Stanford Achievement Test - Form KM, Intermediate Level - Battary Median	74	.84	.84	.91	71	.82	.90	.92

In general, the DAT scores have shown high correlations with achievement tests measuring comparable skills and knowledge.

Reliability: Reliability was studied using the split half technique with the computed correlation coefficients corrected by the Spearman-Brown formula. The VR, NA, and VR+NA coefficients (given separately for form L and M, for boys and girls, and for each grade 8 through 12) range from .83 to .96. The tenth grade values for Form L are: for boys, Verbal Reasoning, r = .93, Numerical Ability, r = .91, VR+NA, r = .95; for girls, Verbal Reasoning, r = .94, Numerical Ability, r = .91, VR+NA, r = .96.

The long term consistency of measurement by the DAT was studied by determining the correlation between 9th grade scores and 12th grade scores for the same set of students studied over the three year period. Verbal Reasoning coefficients of correlation were .87 for boys (N = 71) and .82 for girls (N = 90); Numerical Ability coefficients for these same groups were .75 for boys, .74 for girls. This study utilized DAT - form A.

Norms: Data from national samples of students have been used to establish percentile norms for the DAT. Separate percentile norms tables are given for each form of the test (L and M), for boys and girls, for Fall (first semester) and Spring (second semester) administration, and for each grade 8 through 12. Samples for each table range from N = 1900 to N = 3100.

Correlation to other tests: The DAT correlates well with most standard intelligence tests. Some examples of the coefficients of correlation found between DAT-VR, DAT-NA and DAT-VR+NA scores and various intelligence tests are given in the following table:

TEST	COEFFICIENTS OF CORRELATION							
	BOYS				GIRLS			
	N	VR	NA	VR+NA	N	VR	NA	VR+NA
Lorge-Thorndike intelligence tests (Form A, Level 4) - taken in 11th grade,								
Verbal	58	.70	.60	.72	59	.85	.78	.86
Non-Verbal	58	.61	.57	.64	59	.72	.69	.74
School and College Ability Tests (Form 2A)								
Verbal	71	.82	.57	.78	59	.83	.64	.80
Quantitative	71	.67	.83	.81	59	.77	.82	.85
Total	71	.85	.79	.90	59	.87	.77	.89

REFERENCES

1. Seymour, L. A., and Bingman, R. M., "Development of Views and Preferences - C" a paper presented at the National Association for Research in Science Teaching Convention, March, 1973.

2. "Product Test Information Summary," Annual Report to U. S. Office of Education McREL, Sept. 1, 1971, p. 71.

3. Steele, J., House, E., and Kerins, T. "An Instrument for Assessing Instructional Climate Through Low Inference Student Judgments," AERA Research Journal, May, 1971, p. 449.

4. Ibid., pp. 447-466.

5. Manual for the Comprehensive Final Examination in First-Year Biology, The Psychological Corporation, New York, N.Y. 10017.

6. Steiner, H. Edwin, "A Study of the Relationship Between Teacher Practices and Student Performance of Selected Inquiry Process Behaviors in the Affective Domain in High School Biology Classes," Unpublished Doctoral Dissertation, University of Texas, Austin, 1970.

7. Koos, E. M., "Technical Report No. 1. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, Explorations in Biology," Mid-Continent Regional Educational Laboratory, 1970.

8. Koos, E.M., "Technical Report No. 4. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, Explorations in Biology," Mid-continent Regional Educational Laboratory, 1972.

9. Burmester, M. A., "Behavior involved in the critical aspects of scientific thinking," Science Education, Vol. 36, 1952, p. 259

10. Kaplan, E. H., "The Burmester Test of aspects of scientific thinking as a means of teaching the mechanics of the scientific method," Science Education, Vol. 51, 1967, p. 353

11. Suchman, J. R., The Elementary School Training Program in Scientific Inquiry, Project 216, U. S. Office of Education, University of Illinois, 1962.

12. Taba, H., "Teaching strategy and learning," California Journal of Instructional Improvement, December, 1963.

13. Bingman, R. M., ed., Inquiry Objectives in the Teaching of Biology, Mid-continent Regional Educational Laboratory, Kansas City, Mo., 1969.

14. Koos, E. M., "Technical Report No. 1. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, Explorations in Biology," Mid-continent Regional Educational Laboratory, 1970.

15. Biological Sciences Curriculum Study, *Comprehensive Final Examination, Form K*, New York: The Psychological Corporation, 1965.

16. Bennett, G. K., Seashore, H. G., and Wesman, A. G., *Differential Aptitude Tests*, New York: The Psychological Corporation, 1959.

17. Tiegs, E. W., and Clark, W. W., *California Basic Skills Test*, New York: Houghton Mifflin Company, 1955-56.

18. Lindquise, E. C., Hieronymus, A. N., *et. al*., *Iowa Tests of Basic Skills*, New York: Houghton Mifflin Company, 1955-56.

19. Anderhalter, O. P., Gawkoski, R. S., and O'Brien, J., *Scholastic High School Placement Test*, Bensonville, Illinois: Scholastic Testing Service, 1959.

20. Watson, G., and Glaser, E. M., *Watson-Glaser Critical Thinking Appraisal, Form YM*, New York: Harcourt, Brace and World, 1961.

21. Koos, E. M., "Technical Report No. 1. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, *Explorations in Biology*," Mid-continent Regional Educational Laboratory, 1970.

22. Ibid., p. 15.

23. Kuder, G. F., and Richardson, M. W., "The theory of the estimation of test reliability," *Psychometrika*, Vol. 2, 1937, p. 151.

24. Koos, E. M., "Technical Report No. 1. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, *Explorations in Biology*," Mid-continent Regional Educational Laboratory, 1970.

25. Cronbach, L. J., "Coefficient alpha and the internal structure of tests," *Psychometrika*, Vol. 16, 1951, p. 297.

26. Koos, E. N., "Technical Report No. 2. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, *Explorations in Biology*," Mid-continent Regional Educational Laboratory, 1971.

27. Koos, E. M., "Technical Report No. 4. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, *Explorations in Biology*," Mid-continent Regional Educational Laboratory, 1972.

28. Bingman, Richard, M., Editor, Anderson, John R., Blankenship, Jacob W., Carter, Jack L., Cleaver, Thomas J., Jones, Willard G., Kennedy, Manert, H., Klinckmann, Evelyn, Koutnik, Paul G., Lee, Addison E., Stothart, Jimmy R., *Inquiry Objectives in the Teaching of Biology*, (a joint publication of Biological Sciences Curriculum Study and Mid-continent Regional Educational Laboratory, Sept. 1969), 147 p.

29. *Manual for the Differential Aptitude Tests*, The Psychological Corporation, New York, N.Y. 10017.

DATA ANALYSIS AND INTERPRETATION

Data Processing

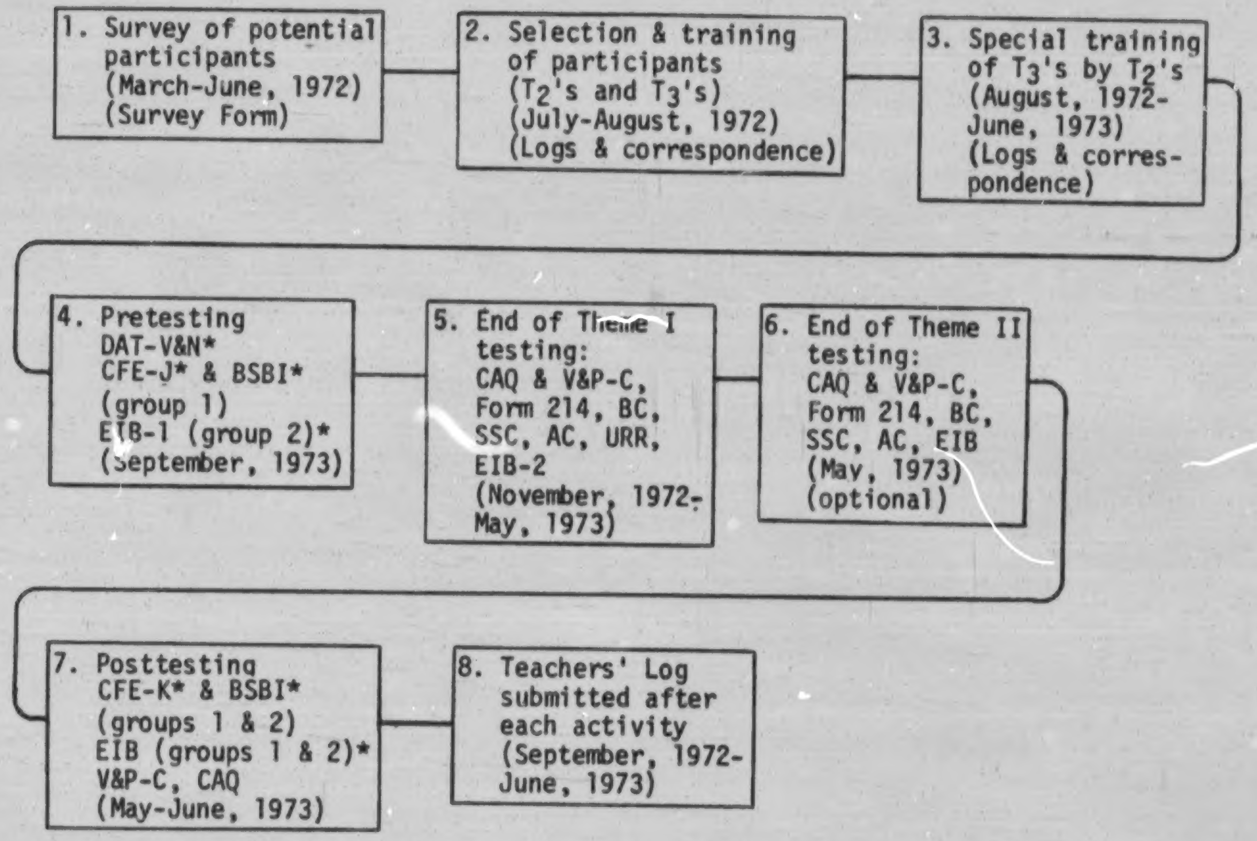
The general sequence of data processing was as follows:

1. Distribution of measuring instruments and instructions to field test participant teachers.
2. Administration of instruments by teachers.
3. Collection of data by McREL.
4. Scanning or key punching data onto cards.
5. Scoring of instruments.
6. Analysis of scores per various groups of subjects.

This basic sequence was repeated three times during the field test to obtain pretest data, interim data after Theme I and posttest data. A brief description of data collected and the approximate times these data were collected are indicated in Chart A.

The statistical processing of data collected during the field test was performed on computers located at the University of Missouri-Columbia (IBM 370/165) and at the University of Kansas (Honeywell 635). For information concerning the particular programs used for the different analyses performed, see Table A. In a few instances, post hoc analyses were computed on desk calculators. All analyses were performed using the student as the sampling unit.

CHART A: Data Collection for IRA Field Test 1972-73


1. Survey of potential participants (March-June, 1972) (Survey Form)
2. Selection & training of participants (T_2's and T_3's) (July-August, 1972) (Logs & correspondence)
3. Special training of T_3's by T_2's (August, 1972-June, 1973) (Logs & correspondence)
4. Pretesting DAT-V&N* CFE-J* & BSBI* (group 1) EIB-1 (group 2)* (September, 1973)
5. End of Theme I testing: CAQ & V&P-C, Form 214, BC, SSC, AC, URR, EIB-2 (November, 1972-May, 1973)
6. End of Theme II testing: CAQ & V&P-C, Form 214, BC, SSC, AC, EIB (May, 1973) (optional)
7. Posttesting CFE-K* & BSBI* (groups 1 & 2) EIB (groups 1 & 2)* V&P-C, CAQ (May-June, 1973)
8. Teachers' Log submitted after each activity (September, 1972-June, 1973)

DAT-V&N = Differential Aptitude Test - Verbal and Numeric
CFE-J&K = Comprehensive Final Examination - Forms J & K
BSBI = Biology Student Behavior Inventory
EIB-1 & 2 = Explorations in Biology - 1A & 1B, and 2A & 2B
CAQ = Class Activities Questionnaire
V&P-C = Views & Preferences - Form C
BC = Biology Content
SSC = Social Skills Checklist
AC = Attitude Checklist
URR = Understanding Role Responsibilities

* Control teachers within the field test design were administered these tests.

TABLE A: Listing of Computer Programs Used for Data Analyses

This PROGRAM was used to obtain this ANALYSIS to support this HYPOTHESIS.*

DATSCOR	SCORED OUTPUT	
BSBSCOR		
EIBSCOR		
PARTPUN		
SORT (U)		
CONDENS		
ANOVAR	Repeated measures analysis of variance	3A; Supplementary Hypothesis 2
MISDATA	Analysis of Variance	3A; Pre-sensitization
BMDO4V	Analysis of Covariance and Newman-Keuls Post Hoc analysis	4A; 4B
SFA41D	Correlations	2
TESTAT	ITEM Analysis	Reliability data for EIB & BSBI
ANOVAR1	Analysis of variance and Newman-Keuls A Posteriori analysis	6
VAPSCOR	Mean, Criterion level classification	1A
SUMCTAB	Descriptive statistics	5

*NOTE: Hypotheses numbers will be revised after whole DATA section is complete; supplementary/EX POST FACTO hypotheses do not now have an identifiable no./title.

Objective 1

Objective 1A: To describe how the program was implemented following each type of teacher training.

Hypothesis 1A: None; this is a descriptive objective.

Objective 1B: To determine whether there is a significant difference in the degree of implementation between groups of teachers receiving different types of training.

Hypothesis 1B: There will be no significant difference between mean implementation ranks (ranking based on the four variables used to describe adequacy of implementation) for teachers receiving different types of training.

Data Analyses/Results: The data related to the adequacy of implementation are shown in Table 1-1. The definitions for type of teacher training and extent of implementation are delineated in the section of this report entitled, OBJECTIVES & NULL HYPOTHESES.

TABLE 1-1: The Adequacy of Implementation of IRA Classes According to Types of Training

TEACHER NO.	TYPE OF TRAINING	PERCENT OF ACTIVITIES COMPLETED BY THEME I	II	III	PERCENT OF STUDENTS REACHING CRITERIA ON ACTIVITIES COMPLETED	EXTENT OF IMPLEMENTATION (W&P-C)** MEAN SCORE	(CAQ)** # OF CATEGORIES EMPHASIZED***	ADEQUACY OF IMPLEMENTATION VA,A,IA	RANK****
20	1a	100	92.8	11	81.8 (19)*	3.96	1,2,3,5,6,7 8,9 (8)	VA	9.5
30	1a	100	85.7	11	77.2 (22)	3.83	1,2,3,4, 5,6,7 (7)	VA	7
40	1a	94.4	92.8	44.4	73.0 (28)	3.67	1,2,5 7, (4)	A	4
01	1b	95.4	0	11	68.5 (16)	3.58	(Not Available)	IA	1
02	1b	68.2	57.1	11	74.7 (10)	3.69	1,2, 6,7,8,9 (6)	A	3
03	1b	100	100	100	74.0 (33)	3.80	1,2,3,4,5 6,7,9 (8)	VA	8
04	1b	100	92.8	44.4	81.2 (33)	4.08	(Not Available)	VA	11
21	2	100	92.8	11	78.5 (19)	3.66	1,2,6, 7,8 (5)	A	5
22	2	100	92.8	11	83.1 (22)	3.80	1,2,3,4,5,6, 7,8,9 (9)	VA	9.5
31	2	100	85.7	11	84.2 (22)	3.65	(Not Available)	A	6
10-14	3	77.3	85.7	11	76.1 (12)	3.60	1,2,3,4, 6,7 (6)	A	2

* The number of activities for which percent of students reaching criteria was reported is noted in parentheses.

** End of Theme I data only used for teachers 20, 01, 03, 22; end of Theme II data only used for teachers 04, 21; average of Theme I and Theme II data used for remaining teachers.

*** 1-application, 2-analysis, 3-synthesis, 4-evaluation, 5-discussion, 6-independence, 7-divergence, 8-ideas valued over grades and 9-enjoyment of ideas; all data from end of Theme II.

**** Rank 1 = lowest implementation; Rank 11 = highest implementation.

Table 1-1 shows that 10 of 11 teachers (or groups of teachers) implemented the program adequately or very adequately--5 were adequate; 5 were very adequate. Only one teacher implemented the program inadequately.

In order to test hypothesis 1B, a method of assigning an implementation rank was developed. The eleven teachers (or teacher group) were ranked on each of the four variables used to describe adequacy of implementation (percent activities completed; percent students reaching criteria; *V&P-C* mean score; number of *CAQ* categories emphasized). Each teacher's mean rank was calculated, and a final implementation rank was assigned based on the mean rank. This implementation rank is given for each teacher (or teacher group) in the last column of Table 1-1. A rank of 1 = lowest implementation; a rank of 11 = highest implementation.

Mean implementation ranks were calculated for each group of teachers (grouping based on type of training). The ranks of the various teachers (or teacher groups) with respect to extent of implementation are recorded in the last column of Table 1--a rank of 1 = lowest implementation; a rank of 11 = highest implementation.

Teachers in Group 1a, who were trained by McREL and also trained other teachers in their districts, had a mean rank of 6.83. Teachers in Group 1b, who were trained by McREL but did not train other teachers, had a mean rank of 5.75. Teachers in Group 2, those trained by Group 1a above, had a mean rank of 6.83. A team teaching group, Group 3, with a team leader who was trained by McREL, had a mean rank of 2.00.

Kruskal and Wallis[1] provided a formula for determining the significance of ranked differences. The formula is:

$$H = \frac{12}{N(N + 1)} \sum \frac{R^2}{n} - 3 (N + 1)$$

Substituting the data:

$$H = 5.86$$

With 3 degrees of freedom this value is not significant at the .05 level.

Interpretation: The lack of statistically significant ranking differences between groups of teachers receiving different types of training, suggests that the various types of training do not result in different extents of IRA program implementation. The data also suggests that teachers (group 2) trained and supervised by a McREL trained teacher-supervisor (group 1a) will not implement the IRA program significantly better than the person who trained them.

In summary, the IRA program was implemented by 14 of 15 teachers in at least an adequate manner. There were no significant differences between types of training with respect to extent of implementation.

Objective 2

Objective 2 - To determine whether there is a relationship between the degree of implementation of the IRA program and student outcomes for biology content achievement, cognitive inquiry skill development, and development of affective qualities of inquiry.

Hypothesis 2 - There is no significant relationship between the three degrees of implementation of the IRA program--very adequate, adequate and inadequate implementation--and the following student outcomes:

1. Comprehensive biology achievement--as measured by the Comprehensive Final Examination-Forms J & K (CFE).
2. Cognitive inquiry skill development--as measured by the Explorations in Biology - Topic 1 (EIB-1).
   - A. Formulate a problem
   - B. Formulate a hypothesis
   - C. Design a study
   - D. Interpret data or findings
   - E. Synthesize knowledge gained from the investigation.
3. Affective qualities of inquiry development--as measured by the Biology Student Behavior Inventory (BSBI). Both the total score and the following sub-scores will be used:
   - A. Curiosity
   - B. Openness
   - C. Satisfaction
   - D. Responsibility

Data Analyses/Results: In order to determine if there were any significant correlations, correlation coefficients were computed between each measure of student outcome and the type of implementation. Student outcome data from all 11 teachers (or teacher groups) given in Table 1-1 were used in this analysis. The type of implementation was determined by the criteria listed for Objective 1. The results of these computations are presented in Table 2-1. Note that EIB-subscale I, formulating problems, was not included since this subscale had previously been shown to be unreliable (see MEASURING INSTRUMENTS section, discussion of EIB-Topic 1 Reliability).

TABLE 2-1: Correlations Between Type of Implementation and Student Outcome Variables

Variable		r	n
EIB III	Formulate Hypotheses	.052	840
EIB IV	Design a Study	.117**	703
EIB V	Interpret Data	.206**	836
EIB VI	Synthesize Knowledge	.120**	814
EIB	Total Score	.168**	703
BSBI A	Curiosity	.101*	593
BSBI B	Openness	.201**	593
BSBI C	Satisfaction	.160**	593
BSBI D	Responsibility	.203**	593
BSBI	Total Score	.227**	593
CFE		.129**	804

*Significant at the .05 level
**Significant at the .01 level

Nine of the correlations are significantly different from zero correlation at the .01 level of significance, and one of them is significant at the .05 level. Although ten of the correlations are significantly different from zero, the correlations are relatively small. Coefficients of determination were calculated for each of the significant correlations as a basis for interpreting the extent to which student outcomes are determined by implementation. These Coefficients of Determination ranged from .0102 (for curiosity) to .052 (for BSBI total score).

Interpretation: The Coefficients of Determination for the significant correlations indicate that at best 5.2 percent of student outcomes are determined by the type of implementation--very adequate, adequate, or inadequate. Thus, for practical purposes, type of implementation does not seem to be substantially related to any of the student outcome variables.

It should be noted that while a meaningful linear relationship between degrees of implementation and student outcomes is not substantiated by these results, there is a significant difference between (a majority of) student outcomes in inadequately implemented classes compared to outcomes in adequately and very adequately implemented classes (discussed in analysis and interpretation section for Objective 4A). The non-meaningful (though statistically significant) relationship demonstrated here apparently can be attributed to the lack of significant differences in student outcome variables between adequately and very adequately implemented classes.

Objective 3A

Objective 3A: To determine whether IRA students, in classes where the program was at least adequately implemented, will show significant increases in biology content, cognitive inquiry skills and affective qualities of inquiry from the beginning of the school year to the end.

Hypothesis 3A: There is no significant gain from pre- to posttesting in biology content knowledge, cognitive inquiry skills and affective qualities of inquiry--as measured by the instruments described in Objective 2--for IRA students in classes where the program was at least adequately implemented.

Data Analyses/Results: In order to determine whether or not there were any significant gains from pretest to posttest for any of the student outcome variables, the repeated measures analysis of variance was computed for each variable. The results of these analyses are presented in Table 3-1. An analysis of variance, non-repeated measures, was also computed; these results are presented in Table 3-2.

Note that this objective and hypothesis dealt only with students in classes where IRA was at least adequately implemented. Therefore, data from teacher 01 were not included in any of these analyses. In addition, since teacher 04 did not pretest students he is not represented in the repeated measures analysis of variance. This teacher was therefore also deleted from the analysis of variance, non-repeated measures.

TABLE 3-1: Number of Students, Pretest and Posttest Means, F Ratios, and Probability Levels for Student Outcome Variables (Repeated Measures Analysis of Variance)

VARIABLE	N	PRETEST MEAN	POSTTEST MEAN	F RATIO	p
EIB III Formulate a Hypothesis	333	10.37	11.87	46.54	.00
EIB IV Design a Study	258	11.47	17.91	536.39	.00
EIB V Interpret Data	331	11.01	23.98	935.37	.00
EIB VI Synthesize Knowledge	309	5.40	7.06	116.88	.00
EIB Total Score	258	39.35	62.36	830.64	.00
BSBI A Curiosity	258	2.61	2.68	2.56	.11
BSBI B Openness	258	3.56	3.66	4.46	.03
BSBI C Satisfaction	258	3.68	3.56	7.46	.01
BSBI D Responsibility	258	3.63	3.71	1.31	.25
BSBI Total Score	258	13.48	13.61	1.00	.32
CFE	290	18.27	19.17	5.74	.02

As can be noted from Table 3-1, 6 of the 12 F ratios are significant beyond the .01 level of significance. Two of the ratios are significant at the .05 level. Thus, there was a significant difference from pretest to posttest on 8 of the 11 student outcome variables. All of the differences were in a positive direction except for the negative difference for the Satisfaction scale on the BSBI. For this scale, there was a significant decrease in the students' scores from pretest to posttest.

Inquiry Role Approach (IRA) students, in classes where at least adequate implementation had occurred, scored significantly (P = less than .01) higher at the end of the school year than at the beginning for the inquiry skill variables: formulating hypotheses, designing studies, interpreting data, synthesizing knowledge and the total score on *Explorations in Biology* (total score includes the foregoing subscales and two items on formulating problems). Thus these cognitive inquiry skills were significantly higher at the end of the school year for IRA students than they were at the beginning of the year. Note that EIB-subscale I, formulating problems, was not studied since this subscale had previously been shown to be unreliable (see MEASURING INSTRUMENTS section, discussion of EIB-Topic 1 reliability).

In the area of affective qualities of inquiry, IRA students showed a significant (P = .03) gain on the openness subscale, a significant (P = .01) decrease on the satisfaction subscale, and no significant difference pre to post on the curiosity and responsibility subscales or BSBI total score.

IRA students showed significant (P = .02) pre to post gain on the biology content variable.

TABLE 3-2: Number of Students, Pretest and Posttest Means, F Ratios, and Probability Levels for Student Outcome Variables (Analysis of Variance, Non-Repeated Measures)

	N		PRETEST	POSTTEST		
VARIABLE	PRE	POST	MEAN	MEAN	F RATIO	P
EIB IA*	568	812	18.93	20.63	36.1	.0000
EIB IB**	573	786	35.86	40.90	113.6	.0000
BSBI A Curiosity	580	519	2.60	2.71	7.53	.006
BSBI B Openness	580	519	3.52	3.70	15.66	.0003
BSBI C Satisfaction	580	519	3.58	3.53	1.74	.18
BSBI D Responsibility	580	519	3.55	3.85	21.57	.0000
BSBI Total Score	580	519	13.26	13.79	17.08	.0002
CFE	589	777	17.56	19.64	40.39	.0000

* EIB-1A is a subscore of EIB 1 which includes EIB subscales I, III, and 12 items from subscale IV.

** EIB-1B is a subscore of EIB 1 which includes 12 additional items from subscale IV and subscales V and VI.

As can be noted from Table 3-2, seven of the eight F ratios are significant beyond the .01 level. All of the differences were in a positive direction. Thus these analyses indicate the null hypothesis (Hypothesis 3A) can be rejected for all variables except BSBI subscale C (satisfaction).

Inquiry Role Approach students, in classes where at least adequate implementation had occurred, scored significantly (P = less than .01) higher at the end of the school than at the beginning for: cognitive inquiry skills as measured by EIB-1A and EIB-1B; affective qualities of inquiry as measured by the BSBI total score and subscale A (Curiosity), B (Openness) and D (Responsibility); and biology content knowledge as measured by the CFE.

The design utilized for testing hypothesis 3A is a quasi-experimental design. Campbell & Stanley[2] have noted that this design may be appropriate in field situations where equivalent or comparable control groups cannot be added. It is further characterized as tending toward superiority in external validity or generalizability over "true" experimental designs. However the most important characteristic of this design for the purposes of this study is its ability to control for the effect of taking a pretest upon the scores of a posttest.

For pretesting, students were randomly distributed into two groups. Group 1 was pretested with the BSBI and CFE instruments; Group 2 was pretested with the EIB-Topic 1 instrument. All students were posttested with all three instruments. Thus Group 2 students acted as a non-pretested control group for the BSBI and CFE instruments; Group 1 students acted as a non-pretested control group for the EIB-1 instrument. An analysis of variance was computed between those students who had the pretest for each variable and those students who did not have the pretest. The results of these analyses are presented in Table 3-3.

TABLE 3-3: Posttest Mean Scores and F Ratios for Comparison of Students With and Without Pretests

TEST	PRETESTED (GROUP 1)		NOT PRETESTED (GROUP 2)			
	MEAN	N	MEAN	N	F	P
EIB III	11.20	399	11.41	432	.67	.58
EIB IV	17.61	339	17.36	361	.95	.67
EIB V	23.96	398	23.24	426	3.40	.06
EIB VI	6.81	387	6.82	415	.02	.89
EIB Total	60.63	339	60.35	361	.11	.74

TEST	PRETESTED (GROUP 2)		NOT PRETESTED (GROUP 1)			
	MEAN	N	MEAN	N	F	P
BSBI A	2.70	214	2.70	378	.02	.89
BSBI B	3.64	214	3.63	378	.02	.89
BSBI C	3.50	214	3.57	378	1.38	.24
BSBI D	3.67	214	3.71	378	.17	.69
BSBI Total	13.51	214	13.60	378	.18	.67
CFE	18.74	406	19.20	393	1.02	.31

The results indicate that there are no significant differences between the two groups on any of the posttest scores. As can be noted in Table 3-3, the means for the two groups are very close and in all cases except for the EIB 3 part score, the group which did not have the pretest scored slightly but not significantly higher than the group of students who had the EIB test as a pretest. For the BSBI scores, the means are again very close and the group of students with the BSBI as a pretest scored slightly but not significantly higher on two of the part scores and the total score. As noted above, none of these differences were significant at the .05 level of significance. For the CFE, the group of students who had the CFE as a pretest scored about half a point higher than the group of students who did not have this test as a pretest, but again the difference was not significant at the .05 level.

It should be noted that the design used here does not control for maturation--pre to post changes resulting from the passage of time rather than treatment. However, a modified Solomon Four-Group Design was used for Objective 6 and posttest only analyses were performed comparing experimental and control groups.

Interpretation:

Repeated Measures Analysis of Variance: Apparently the students in the IRA program strongly developed cognitive inquiry skills as measured by the EIB-1 instrument. Significant development of the knowledge of biology, as measured by the CFE, is also indicated.

While the affective quality of openness was significantly increased, total BSBI scores indicate that overall development of affective qualities of inquiry was not significant. (However it is important to note that IRA students did show change in a positive direction on the BSBI total score.) Scores on the BSBI subscale for satisfaction showed a significant decrease.

Analysis of Variance, Non-Repeated Measures: These results indicate IRA students strongly developed cognitive inquiry skills as measured by the EIB-1 instrument, biology content knowledge as measured by the CFE, and affective qualities of inquiry (with the exception of Satisfaction) as measured by the BSBI instrument.

Comparison of Analyses: When the results of the repeated measures analysis of variance and analysis of variance (non-repeated measures) are compared, the following should be noted: 1) Cognitive inquiry scores are shown to be significantly higher on the posttest than on the pretest. The probability levels do change somewhat from one analysis to the other, but all P values are well below the .01 level. 2) Biology content knowledge scores are significantly higher on the posttest than on the pretest. The probability level changes from .02 when the repeated measures analysis is used to 1.3×10^{-6} when the non-repeated measures analysis is used. 3) The results regarding the affective qualities of inquiry are substantially different for the two analyses. Subscale A (Curiosity) shows a non-significant gain pre-to-post using repeated measures analysis of variance (ANOV), but shows a significant gain using non-repeating measures ANOV. Subscale B (Openness) shows a significant gain using both analyses, but the probability level changes from .03 (repeated measures ANOV) to 2.5×10^{-4} (non-repeated measures ANOV). Subscale C shows a significant (P = .01) loss pre-to-post using repeated measures ANOV, but shows a small, non-significant loss using non-repeated measures ANOV. Subscale D and total BSBI scores change as did subscale A, from non-significant gains (using repeated measures) to significant gains (using non-repeated measures). In view of the varying results, particularly for the measures of affective qualities, some of the limitations of each techniques are noted.

The repeated measures ANOV technique requires each student included to have both a pre and psot test for the variable being analyzed. Each subject serves as his own control, and a source of unexplained variation is accounted for. At the same time a smaller unique sample results; the possibility of an atypical sample is increased. In fact, the atypical nature may be enhanced by the fact that the students included have been conscientious enough to have taken both a pre and posttest. (Students who drop from the course, enter after pretesting, erroneously complete the test so that data is rejected, etc., are eliminated.)

Use of the non-repeated measures ANOV technique allows for a larger and therefore probably more representative sample (note the comparison of N's between Tables 3-1 and 3-2). However there is more uncontrolled variance since each student cannot act as his own control.

In view of the pre-to-post gains shown for these subscales (subscale B gains significant at the .03 level) and the total BSBI score using the repeated measures analysis of variance, coupled with the significant pre-to-post gains shown for these three subscales and the total BSBI score using the non-repeated measures analysis of variance, and the superior results of the experimental group over the control group as discussed in Objective 6, it is apparent that students in the Inquiry Role Approach program do develop the affective qualities of inquiry as measured by the BSBI instrument.

Pretest Sensitization: It also appears that there is little pretest sensitization operating on the posttest scores for the instruments used in this study.

Note that student pretest mean scores for cognitive inquiry (EIB subscales III through VI and total score on Table 3-1; EIB 1A and 1B scores on Table 3-2 are near the chance level as expected (e.g., EIB total score, chance = 40.97; EIB total score, Table 3-1, pretest = 39.35). This supports the assumption that the instrument was properly administered by teachers and seriously responded to by students.

Subscale scores on the BSBI have a possible range from 1 to 5; scores at or below 3 are considered neutral. That is, they indicate an indifference to the affective quality being measured. Note that both pretest and posttest mean scores for subscale A (Tables 3-1 and 3-2) indicate that curiosity, as measured by the BSBI, is not at a meaningful level. Means on all other subscales and the total score are well above the "neutral" level.

Objective 3B

Objective 3B: To determine whether IRA students, in classes where the program was at least adequately implemented, prefer the social behaviors, cognitive behaviors, and classroom procedures characteristic of the IRA program.

Hypothesis 3B: A majority of the teachers--having performed at least adequate implementation--will not report a class mean score of greater than 3.50 for the preference items on the instrument Views and Preferences - Form C. (A mean score of greater than 3.50 indicates more than 50 percent of the students prefer the set of social behaviors, cognitive behaviors and classroom procedures presented in the instrument.)

Data Analysis/Results: A description of the scoring of the instrument Views and Preferences - Form C is found in the section MEASURING INSTRUMENTS. A student mean score of greater than 3.50 indicates that on the average over 50 percent of the students responding prefer the behaviors and procedures presented in the instrument. The student mean scores for preference items are presented in Table 3-3. This table includes data both from end of Theme I (approximately mid-year) and end of Theme II (near end of year for most teachers). This objective included only students in classes where the IRA program was at least adequately implemented. Therefore data from teacher 01 are not included in Table 3-3.

TABLE 3-3: V&P-C Preference Items Mean Scores of IRA students in Adequately and Very Adequately Implemented Classes

TEACHER NO.	TYPE OF IMPLEMENTATION	MEAN SCORE SOCIAL BEHAVIORS* THEME I	THEME II	MEAN SCORE COGNITIVE BEHAVIORS* THEME I	THEME II	MEAN SCORE CLASS PROCEDURES* THEME I	THEME II	MEAN SCORE TOTAL-ALL PREFERENCE ITEMS* THEME I	THEME II
20	A	3.967	NA**	3.867	NA	3.767	NA	3.888	NA
30	A	3.685	3.954	3.180	3.524	3.804	3.692	524	3.740
40	A	3.887	3.858	3.664	3.47	3.875	3.625	3.802	3.670
02	A	3.976	NA	3.445	NA	3.933	NA	3.771	NA
03	VA	4.019	3.986	3.414	3.500	4.000	3.885	3.792	3.774
04	VA	NA	4.156	NA	3.888	NA	4.268	NA	4.080
21	A	NA	3.947	NA	3.374	NA	3.846	NA	3.715
22	VA	3.979	NA	3.488	NA	3.823	NA	3.765	NA
31	VA	3.917	3.854	3.655	3.319	3.927	3.792	3.822	3.643
10-14	A	3.821	3.754	3.325	3.433	3.715	3.883	3.616	3.671
ALL TEACHERS REPORTING		3.906	3.930	3.505	3.501	3.856	3.856	3.748	3.756

* Number of preference items for social behaviors = 8, for cognitive behavior = 7, for class procedures = 4, for all preference items = 19.

** NA = instrument not administered.

As seen in Table 3-3, all mean scores reported for social behaviors and class procedures were above 3.50. In addition, all mean scores reported for the total score of all preference items were above 3.50. Three of eight teachers reporting at end of Theme I and three of seven teachers reporting at end of Theme II reported mean scores for cognitive behaviors above 3.50.

Interpretation: With respect to the social behaviors and class procedures characteristic of the IRA methodology, students demonstrated strong preferences in classes where the program was at least adequately implemented. For the cognitive behaviors characteristic of IRA, approximately half of the students demonstrated a preference. The mean total scores for all preference items on the V&P-C indicated a general preference by IRA students for the behaviors and procedures characteristic of IRA. Thus the null hypothesis (Hypothesis 3B) is rejected.

Objective 4

Objective 4: To determine whether there are significant differences in student outcomes in biology content knowledge, cognitive inquiry skills and affective qualities of inquiry between students in the following subgroups:

1. Students in classes where the program was inadequately implemented, adequately implemented, and very adequately implemented.
2. Students with verbal and numerical ability at the 75th percentile or above, from the 50th to the 74th percentile, from the 25th to the 49th percentile, and at the 24th percentile or below.

Hypothesis 4A: There is no significant difference in student outcomes--biology content knowledge, cognitive inquiry skills and affective qualities of inquiry as measured by the instruments described in Objective 2--for students in classes with different degrees of implementation of the IRA program.

Data Analysis/Results: In order to test Hypothesis 4A, the analysis of covariance was computed for each of the eleven student outcome variables (note that EIB-subscale I was not used due to subscale unreliability as discussed previously). Pretest scores were held constant for each variable analyzed. The Newman-Keuls statistical test was used to determine which pairwise differences were significant.

Table 4-1 presents the adjusted posttest means and F ratios for comparing student outcome variables for the three subgroups based on degree of implementation. Table 4-2 presents the results of the Newman-Keuls analysis.

TABLE 4-1: Adjusted Means and F Ratios for Comparing Subgroups Based on Degree of Implementation

VARIABLE	INADEQUATE IMPLEMENTATION		ADEQUATE IMPLEMENTATION		VERY ADEQUATE IMPLEMENTATION		F RATIO	DF
	ADJUSTED MEAN	N	ADJUSTED MEAN	N	ADJUSTED MEAN	N		
EIB III	9.48	22	11.69	204	12.11	129	5.81**	(2,351)
EIB IV	14.75	23	17.95	144	18.24	114	11.62**	(2,277)
EIB V	18.52	25	23.89	202	25.33	129	13.46**	(2,352)
EIB VI	6.35	25	6.96	180	7.57	117	5.57**	(2,330)
EIB Total	49.14	23	62.35	149	64.38	114	21.99**	(2,277)
BSBI A	2.79	58	2.65	144	2.76	114	1.84*	(2,312)
BSBI B	3.67	58	3.64	144	3.70	114	.30	(2,312)
BSBI C	3.54	58	3.51	144	3.61	114	.92	(2,312)
BSBI D	3.65	58	3.76	144	3.78	114	.481	(2,312)
BSBI Total	13.71	58	13.56	144	13.80	114	.493	(2,312)
CFE	20.20	59	19.17	197	20.38	93	1.68*	(2,345)

*Sig. at the .25 level.

**Sig. at the .01 level.

TABLE 4-2: Newman-Keuls Post Hoc Analysis for Extent of Implementation

	IA/A	IA/VA	A/VA
EIB III	**	**	
EIB IV	**	**	
EIB V	**	**	
EIB VI		**	
EIB Total	**	**	

* Significant at .05 level
** Significant at .01 level

Five of the eleven F ratios are significant at the .01 level of significance. These involved the following student outcome variables: EIB III, EIB IV, EIB V, EIB VI, and EIB total score.

For the EIB III comparisons, the Newman-Keuls post hoc analysis indicated that the achievement level of the students under the teacher with inadequate implementation was significantly below both the other subgroups. For the EIB IV scores, the post hoc test indicated that the students under the inadequate implementation teacher were significantly lower than both the other subgroups. The same pattern is true for the EIB V and EIB total score. For EIB VI only the very adequate and inadequate means were significantly different. All of the comparisons were significant at the .01 level of significance.

Interpretation: The data presented suggests that at least adequate implementation is necessary to attain development of cognitive inquiry, but not necessary for development of affective qualities and biology content knowledge. Much caution must be exercised in interpreting this data. Data from only one teacher is included in the "inadequate implementation" category. Further, the students in this teacher's classes were all ninth grade students (compared to primarily tenth grade students in adequately and very adequately implemented classes); and students were in class only 180 minutes/week. This teacher strongly emphasized social and attitudinal development (note that there was no significant difference between this teacher's class and all other classes in the area of affective qualities). This emphasis contributed to the lack of use of much of the IRA programs materials (no activities in Theme II and only 11 percent of Theme III activities were completed). The lack of completion of IRA activities may have strongly contributed to the significantly lower cognitive inquiry scores. Further studies using more carefully controlled groups (in terms of grade level, class structure, etc.) and larger sample size might give more conclusive results.

The question is also raised as to the relative validity of the four variables used to evaluate degree of implementation. It may be appropriate to place greater emphasis on certain variables (for example, percent of IRA activities completed) than on others.

Hypothesis 4B: There is no significant difference in student outcomes--biology content knowledge, cognitive inquiry skills and affective qualities of inquiry as measured by the instruments described in Objective 2--for students with different verbal and numerical abilities. (Verbal and numerical abilities: Students will be stratified according to their Differential Aptitude Test pre-test scores into four verbal and four numerical ability groups for tenth grade students only--75th percentile and above, 50th to 74th percentile, 25th to 49th percentile, and 24th percentile and below.)

Data Analyses/results: The analysis of covariance was also used to determine whether or not there were any significant differences in student outcome variables (EIB-subscale I not included) among the four subgroups based on both the DAT-Verbal and the DAT-Numerical scores. Pretest scores were held constant for each variable analyzed. The results of these analyses are presented in Tables 4-3 and 4-4 for, respectively, the subgroups based on DAT-Verbal scores and DAT-Numerical scores. The Newman-Keuls analysis was used to determine which pairwise differences were significant; these results are presented in Tables 4-5 and 4-6.

Data used in these analyses include data for students of all eleven teachers (or teacher groups) listed on Table 1-1 (see Data Analysis/Results for Objective 1A). With two exceptions (teachers 03 and 04) teachers administered the DAT-Verbal and DAT-Numerical instruments to students within the first week of the school year.

Teacher 03 did not administer DAT in the fall of the year. Students at her school were scheduled to take the DAT tests during the year as part of a school-wide testing program. It was planned to utilize test scores from this testing. However, it was found that the tests administered were new forms of the DAT being developed by The Psychological Corporation, Forms S and T. Verbal and Numerical raw scores and percentile rankings were available for the majority of Teacher o3's students from testing with Forms S and T. In order to convert these scores into equivalent DAT Form L Verbal and Numerical raw scores for fall testing, the Verbal and Numerical percentile were considered equal to the percentile each student would have attained on the DAT Form L. Using tables on pages 3-10, 3-11, and 3-12 of the Manual for the Differential Aptitude Tests raw scores were assigned.

Teacher 04 did not administer DAT in the fall of the year. DAT Form L scores were available from the school district records for this teacher's students. However, it was found that these scores were from tests administered two years previous, in the students' eighth grade. Since these raw scores would not reflect student abilities at the present time, the Verbal and Numerical percentiles were used to convert to a current (tenth grade) raw score, again using the appropriate table (p. 3-10) in the Manual for the Differential Aptitude Tests.

TABLE 4-3: Adjusted Means and F Ratios for Comparing Subgroups Based on Quartiles on DAT Verbal Scores

VARIABLE	FIRST QUARTILE ADJUSTED MEAN	N	SECOND QUARTILE ADJUSTED MEAN	N	THIRD QUARTILE ADJUSTED MEAN	N	FOURTH QUARTILE ADJUSTED MEAN	N	F RATIO
EIB 3	10.19	35	10.07	51	12.02	82	12.78	117	13.06**
EIB 4	15.69	27	16.83	33	17.66	66	19.00	90	10.54**
EIB 5	18.89	38	20.26	42	23.51	93	27.27	109	40.25**
EIB 6	5.80	34	6.87	38	6.76	88	7.61	106	9.67**
EIB Total	53.59	27	54.45	33	60.79	66	68.17	90	34.38**
BSBI A	2.54	21	2.52	48	2.67	65	2.77	82	1.90
BSBI B	3.45	21	3.43	48	3.60	65	3.76	82	3.05*
BSBI C	3.38	21	3.37	48	3.49	65	3.65	82	2.98*
BSBI D	2.91	21	3.35	48	3.72	65	3.90	82	6.52*
BSBI Total	12.46	21	12.76	48	13.51	65	13.94	82	5.26*
CFE	16.11	25	18.41	45	17.27	72	19.89	98	3.68*

* Significant at .05 level
** Significant at .01 level

TABLE 4-4: Adjusted Means and F Ratios for Comparing Subgroups Based on Quartiles on DAT Numeric Scores

VARIABLE	FIRST QUARTILE ADJUSTED MEAN	N	SECOND QUARTILE ADJUSTED MEAN	N	THIRD QUARTILE ADJUSTED MEAN	N	FOURTH QUARTILE ADJUSTED MEAN	N	F RATIO
EIB 3	11.11	86	11.72	67	12.01	96	13.06	21	2.41
EIB 4	16.18	61	17.32	52	18.79	81	19.69	19	13.03**
EIB 5	21.09	85	22.99	75	25.96	94	28.13	22	17.81**
EIB 6	6.63	76	6.82	72	7.29	41	7.96	20	4.00*
EIB Total	56.41	61	60.26	52	65.22	81	69.26	19	14.64**
BSBI A	2.44	52	2.66	68	2.74	78	2.85	17	3.36*
BSBI B	3.46	52	3.59	68	3.72	78	3.52	17	1.97
BSBI C	3.40	52	3.51	68	3.53	78	3.65	17	.94
BSBI D	3.21	52	3.63	68	3.85	78	3.80	17	4.51*
BSBI Total	12.58	52	13.41	68	13.81	78	13.65	17	4.45*
CFE	16.21	67	17.80	79	19.82	75	21.62	18	6.35*

* Significant at .05 level
** Significant at .01 level

TABLE 4-5: Newman-Keuls Post Hoc Analysis for DAT-Verbal Quartiles

	1ST≠/2ND	1ST/3RD	1ST/4TH	2ND/3RD	2ND/4TH	3RD/4TH
EIB 3		**	**	**	**	
EIB 4	**	*	*		*	
EIB 5	**	**	**	**	**	
EIB 6			**		**	**
EIB Total	**	**	**	**	**	
BSBI B		*				
BSBI C						
BSBI D		*	**		**	
BSBI Total		*	**		*	
CFE			*			

* Significant at .05 level
** Significant at .01 level

TABLE 4-6: Newman-Keuls Post Hoc Analysis for DAT-Numerical Quartiles

	1ST/2ND	1ST/3RD	1ST/4TH	2ND/3RD	2ND/4TH	3RD/4TH
EIB 4		*	*	**	*	
EIB 5		**	**	**	**	
EIB 6		*	**			
EIB Total		**	**	*	**	
BSBI D			*		*	
BSBI Total						
CFE		*	**			

* Significant at .05 level
** Significant at .01 level

≠ Key: 1ST = 4TH quartile
2ND = 3RD "
3RD = 2ND "
4TH = 1ST "

As indicated in Tables 4-3 and 4-4, all but one of the F ratios for the total and four subscale scores on the EIB are significant; all but three of the F ratios for the total and four subscale scores on the BSBI are significant; and the F ratios for the CFE are significant. Tables 4-5 and 4-6 indicate which of the pairwise comparisons are significant. It should be noted that, although the F ratios were significant for BSBI-subscale C compared to DAT-Verbal and BSBI-total score compared to DAT-Numerical, the Newman-Keuls analysis did not result in any significant pairwise differences.

In order to further clarify the possible relationships between student outcome variables and DAT scores, correlation coefficients were computed between each measure of student outcome and the DAT scores. Table 4-7 presents the results of this analyses.

TABLE 4-7: Correlations Between Posttest Student Outcome Variables and DAT-Verbal and DAT-Numerical Scores

	r* DAT-V	N	r* DAT-N	N
EIB 3	.417	742	.318	718
EIB 4	.450	636	.425	623
EIB 5	.550	735	.463	722
EIB 6	.361	716	.278	703
EIB Total	.610	636	.525	623
BSBI A	.242	522	.249	499
BSBI B	.484	522	.435	499
BSBI C	.299	522	.294	499
BSBI D	.462	522	.428	499
BSBI Total	.507	522	.479	499
CFE	.481	717	.482	687

* All correlations are significant at 0.01

All of the correlations given in Table 4-7 are significant at the .01 level.

Interpretation: It is apparent from the Newman-Keuls test results shown in Tables 4-5 and 4-6 that student outcomes in cognitive inquiry as measured by the instrument EIB-Topic 1 are related to both DAT-Verbal and Numerical scores since there are a number of significant differences between the various quartile subgroups. The correlation coefficients for EIB-Total scores (the coefficients indicating significant positive linear relationships) also support this view. It should be noted that these coefficients (r = .610, DAT-Verbal-EIB total score; r = .525, DAT-Numerical-EIB total score) are near the value previously reported by Koos[3] (Tech Report #1, 1970) (r = .63, DAT-Numerical + Verbal Composite score - EIB-1 total score).

Tables 4-5 and 4-6 also indicate that student outcomes for affective qualities measured by the BSBI are related to DAT-Verbal scores. Only two pairwise comparisons for BSBI - subscale D show significant differences; BSBI-total scores show no significant differences in pairwise comparisons. Therefore there does not appear to be a substantial relationship between BSBI and DAT-Numerical. The correlation coefficient (.479) would support this view. This is as expected since the BSBI instrument is designed to measure affective qualities.

In the comparison of CFE to DAT-Verbal, only one quartile pairing, 1st to 4th, shows a significant difference (p = .05). Two pairings show significant difference when CFE and DAT-Numerical or compared (1st to 3rd, p = .05; 1st to 4th, p = .01). CFE and DAT scores therefore are apparently related, but not to the degree shown for EIB and DAT scores. This view is again supported by the correlation coefficients (r = .481, CFE-DAT-Verbal; r = .482, CFE-DAT-Numerical).

In summary, student outcomes are generally related to students' verbal and numerical ability as measured by the DAT; the null hypothesis (Hypothesis 4B) is rejected. The highest degree of relationship to verbal/numerical ability is shown for cognitive inquiry skill development as measured by the EIB. These results are generally as expected. The data reported here in Tables 4-3 and 4-4 may also be helpful to future teachers using the IRA program for evaluating student performance.

Objective 5

Objective 5: To determine whether IRA students will demonstrate criterion level performance in biology content knowledge, cognitive inquiry skills, social skills and affective qualities of inquiry at an interim point in the program.

Hypothesis 5: IRA students will not demonstrate the following criteria levels when tested at the end of Theme I:

	CRITERION SCORE	PERCENT OF TOTAL SCORE	CHANCE SCORE
1. Theme I biology achievement--as measured by a 72-item biology content test.*			
A. Information and definition items:	12.5	50%	6.25
B. Application and inquiry process items:	23.5	50%	11.75
2. Theme I cognitive inquiry skill development--as measured by EIB-2A & 2B.			
A. Formulate a problem:	1.7	85%	1.6
B. Search for information:	20.9	55%	19.5
C. Formulate a hypothesis:	9.35	55%	7.5
D. Design a study:	28.6	55%	20.4
E. Interpret data or findings:	17.6	55%	16.0
F. Synthesize knowledge gained from the investigation:	9.9	55%	8.6
3. Theme I social skill development--as measured by the Social Skills Checklist* and Understanding Role Responsibilities* quiz.			
A. Understanding Role Responsibilities:	30	75%	10
B. Social Skills Checklist:	28	56%	

---

*These instruments are found in Inquiry Role Approach THEME I MANUAL, Activity 121.

	CRITERION SCORE	PERCENT OF TOTAL SCORE	CHANCE SCORE
4. Theme I affective qualities of inquiry--as measured by the Attitude Checklist.*			
A. Attitude Checklist:	33	51%	

Data Analyses/Results: Interim data (data from the end of Theme I assessment) was recorded by teachers on summary sheets. A separate summary sheet, showing scores for each of the instruments and subscales given in hypothesis 5 above, was completed for each student. When summary sheets were received, a sample of 20 to 30 from each teacher was randomly selected for analysis. Using this sample, mean scores were calculated for each instrument and/or subscale.

The sample for teacher group 10 (teachers 10 through 14) was larger (approximately 80) since the sample represents a large number of students. Teacher 40 did not send summary sheets but did send mean scores calculated for all 103 students who took the Theme I assessment.

Prior to receipt of the data, criterion levels were set for each instrument and subscale. Criterion levels were the cooperative professional judgment of three program development specialists, one research and evaluation specialist, and two experienced IRA teachers (four years teaching IRA).

A summary of the data from the end of Theme I assessment is given in Table 5-1.

---

*These instruments are found in Inquiry Role Approach THEME I MANUAL, Activity 121.

TABLE 5-1: End of Theme I Achievement of IRA Students

TEACHER NO:	01		02		03		04		10-14		20		21		22		30		31		40	
*SCALE NO.	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N
1	13.0	24	14.7	23	14.1	22	--	--	--	--	14.8	20	17.1	20	14.1	26	14.9	30	12.5	25	--	--
2	26.6	24	27.3	23	21.1	22	--	--	--	--	28.9	20	28.3	20	25.5	26	28.6	30	25.1	25	**47	103
3	20.9	24	20.8	24	18.9	21	30.0	25	20.0	85	20.2	20	29.8	20	26.7	26	29.3	28	22.4	25	28	103
4	33.8	24	43.2	24	40.2	22	49.4	26	33.4	77	35.9	20	41.0	20	42.4	26	38.8	26	43.2	25	42	103
5	43.7	24	48.4	23	46.6	22	--	--	41.7	82	44.4	20	45.6	20	49.4	26	43.0	27	51.1	25	47	103
6	1.83	24	2.0	22	2.0	22	--	--	--	--	1.6	20	1.79	19	1.77	26	1.33	27	1.76	25	--	--
7	5.75	24	7.5	22	6.8	21	--	--	--	--	7.2	19	10.8	20	6.5	20	6.4	21	8.24	25	--	--
8	22.3	24	24.8	23	22.4	22	--	--	--	--	24.4	20	25.1	20	22.4	24	25.2	25	21.4	25	--	--
9	26.1	24	26.8	23	29.0	21	--	--	--	--	30.1	20	30.3	19	20.4	25	30.0	26	23.9	25	--	--
10	19.9	24	20.6	22	21.4	21	--	--	--	--	21.7	20	20.8	20	20.3	21	22.0	25	19.9	25	--	--
11	10.4	24	10.6	22	11.4	21	--	--	--	--	12.0	20	9.7	19	11.1	22	11.2	20	9.58	25	--	--
LEIB	--	--	--	--	--	--	***3.33	30	---	--	--	--	--	--	--	--	--	--	--	--	--	--
EIB Total (Scales 6 + 7 + 8 + 9 + 10 + 11)	86.3	24	92.3	22	93.0	21	--	--	90.9	72	97.0	19	99.1	19	82.5	20	96.1	20	84.9	25	100	103

*Scale No.	Description	Maximum Possible Score	Criterion Score	Percent of Maximum Score	Chance Score
1	Biology content (information and definition)	25	12.5	50%	6.25
2	Biology content (application and inquiry process)	47	23.5	50%	11.75
3	Understanding Role Responsibilities	40	30.0	75%	10.0
4	Social Skills Checklist	55	28.0	56%	--
5	Attitude Checklist	65	33.0	51%	--
6	Formulating Problems	2	1.75	85%	1.6
7	Searching for Information	17	9.35	55%	7.5
8	Formulating Hypotheses	38	20.9	55%	19.5
9	Designing a Study	52	28.6	55%	20.4
10	Interpreting Data or Findings	32	17.6	55%	16.0
11	Synthesizing Knowledge	18	9.9	55%	8.5
	Explorations in Biology Total (Scale 6 + 7 + 8 + 9 + 10 + 11 = EIB Total)	159	88.1	55%	73.5

**Teacher 40 reported a scale 1 + scale 2 total score; the mean total score reported was 47 which meets the combined criterion level (scale 1 criterion, 12.5 + scale 2 criterion, 23.5 = 36.0).

***A = 4.0, B = 3.0, C = 2.0, etc., thus a 3.33 or B+ average was maintained on the Laboratory Explorations in Biology (end of Theme I) report.

Scales 1 & 2: A criterion level of 12.5 was set for scale 1, Biology Content - Information and Definition, a 25-item scale with a chance score of 6.25. A criterion level of 23.5 was set for scale 2, Biology Content - Application and Inquiry Process, a 47-item scale with a chance score of 11.75. These criterion were based on the Manual for the Comprehensive Final Examination in First-Year Biology,[4] the BSCS Quarterly Achievement Tests - Yellow Version Manual,[5] and The Resource Book of Test Items for Biological Science - An Inquiry Into Life, 2nd Edition[6] from which the biology content items were taken. From these sources it was determined that a national sample of students with mean scores of 50 percent on tests with norms established (e.g., CFE) were at about the 40th percentile in the norm population. Since the norm population had DAT-Verbal + Numerical composite score mean at the 70th percentile, and the field test population mean was near the 60th percentile on DAT-Verbal, a mean score of 50 percent on the biology content seemed roughly comparable to an expected level of achievement for students at this level.

Scale 3: Understanding Role Responsibilities is a 20-item instrument with a maximum score of 40, chance score of 10, and a criterion of 30. This criterion level was based solely on the professional judgment of staff members as stated in above group. Of the 11 teachers (or teacher groups) reporting, only 1 reported meeting this level, and the report indicated this level was reached after 2 to 3 administrations of the instrument. Mean scores ranged from 18.9 to 30. While the criterion level was not met, it should be noted that criterion was high and that all teachers reported student mean scores well above the chance level.

Scale 4: The Social Skills Checklist is an 11-item instrument with a criterion level of 28, a maximum of 55, and no chance score. Scores were to be interpreted as follows:

excellent	50 - 55
good	39 - 49
satisfactory	28 - 38
poor	17 - 27
very poor	11 - 16

The mean scores ranged from 33.4 to 49.4 for the 11 reporting teachers (or groups); thus all exceeded the criterion level. Three reported a mean score between 28 and 38 for "satisfactory" social skills ability, and eight reported mean scores in the "good" category.

Scale 5: The Attitude Checklist is a 13-item instrument with a criterion level of 33, a maximum score of 65, and no chance score. Scores were to be interpreted as follows:

very frequently	59 - 65
frequently	46 - 58
sometimes	33 - 45
seldom	20 - 32
rarely or never	13 - 19

The criterion level of 33 was exceeded by all 10 reporting teachers. The mean scores ranged from 41.7 to 51.1. Four teachers reported scores in the "sometimes" range, and six reported scores in the "frequently" range.

Scales 6 through 11: The total mean scores on the Explorations in Biology-Topic 2 have ranged from 53 to 109[7] when administered to secondary and college students at various times of the year. Because of this wide range and limited description of subjects, time of testing, and instruction received, the professional judgment of the above mentioned staff was used for establishing criteria levels for the various parts and total scores of EIB-2.

A criterion level of 55 percent was arbitrarily established on the total and all sub-parts, except Formulating Problems which was set at 87.5 percent because of the high chance score. This level, 55 percent, seemed to be a meaningful level because it was well above the overall chance percentage of 46.3 percent and all EIB's are considered difficult tests.

The reported mean scores for scale 6, Formulating Problems, ranged from 1.33 to 2.0. Six of the eight reporting teachers reported mean scores exceeding the 1.75 criterion level.

Scale 7, Searching for Information, had a criterion score of 9.35. The reported means ranged from 6.4 to 10.8; only one of eight exceeded the criterion level.

Scale 8, Formulating Hypotheses, had a criterion score of 20.9. The eight reported mean scores ranged from 21.4 to 25.7; all exceeded criterion.

The criterion level for scale 9, Designing a Study, was 28.6. The reported mean scores ranged from 20.4 to 30.3. Four of the eight exceeded criterion.

The criterion level for scale 10, Interpreting Data or Findings, was 17.6. The reported mean scores ranged from 19.9 to 22.0. Thus all eight exceeded criterion.

Finally scale 11, Synthesizing Knowledge, had a criterion score of 9.9. The reported means ranged from 9.58 to 12.0; six of the eight exceeded criterion.

The EIB-2 total score criterion level was 88.1 (55.4% of the 159 total possible points). Teacher group 10 (teachers 10-14) and teacher 40 did not report scores for scales 6 through 11, but did report mean scores for EIB-2 total score. Including these results, EIB-2 total score means ranged from 82.5 to 100.0, and 7 of 10 teachers reported means exceeding the criterion level. The three means below criterion were nonetheless well above chance.

Teacher 04 reported on the Laboratory Exploration in Biology (LEIB) instead of the Exploration in Biology (EIB), and a mean score of 3.33 was indicated. This was equivalent to a B+ average.

Interpretation:

Scales 1 & 2: Since all teachers reporting met criteria for scale 1, and all but one teacher met criteria for scale 2, one can conclude that the biology content achievement for Theme I was adequate.

Scale 3: In terms of the criterion level stated for scale 3, student understanding of role responsibilities at the end of Theme I appears to be inadequate. Recall, however, that criteria were not empirically based and may have been set too high. This view would be supported by the data noted above--all student means were well above the chance level even though most did not meet criterion. Also, by the end of Theme I most students had only experienced the performance of one role; therefore, it may be likely that they had adequate knowledge of their own role but inadequate knowledge of the other three.

These results indicate a need to review the suggested criterion level. They also suggest the possible usefulness in further study of the function of actual role performance in learning role theory.

Scale 4: Since all reported student means exceeded criterion, it is suggested that student social skills ability at the end of Theme I was at least satisfactory. Caution must be exercised, however, since this is a checklist which utilizes student ratings which may not be highly valid or reliable. (See MEASURING INSTRUMENTS section for full discussion of this instrument.)

Scale 5: As with scale 4, the reported student mean all exceeded criterion and suggest that students adequately demonstrate behaviors which are indicative of the attitudes Theme I is intended to develop. Again, however, caution must be exercised due to the nature of the instrument. (See MEASURING INSTRUMENTS section for full discussion of this instrument.)

Scales 6-11: Student means reported for two scales of the EIB-2 instrument--scale 7, Searching for Information, and scale 9, Designing a Study--did not show adequate development at end of Theme I. However, the remaining scales, 6, 8, 10, 11, did show criterion level and above mean scores. And the overall EIB-2 criterion was met by students in classes of 7 of the 10 reporting teachers. Therefore adequate development of cognitive inquiry skills by the end of Theme I is indicated.

It should be noted that each theme in the IRA program stresses the development of different areas of inquiry. (The thematic structure of IRA is discussed more fully in the PROGRAM DESCRIPTION section.) Theme I is designed to particularly develop problem and hypothesis formulation. While problem formulation is almost impossible to evaluate reliably with only two items on an instrument (note that maximum score for scale 6 is 2.0), the hypothesis formulation scale, scale 8, should be more carefully viewed. Note that all eight reporting teachers reported student mean scores exceeding criterion. Further note that searching for information (scale 7) and designing a study (scale 9) are the two areas of inquiry emphasized in Theme II of IRA. Therefore, the below criterion results reported for scales 7 and 9 are not highly valid indicators of the degree of cognitive inquiry skill development in Theme I.

In summary, biology content knowledge, social skills development, development of attitudinal qualities of inquiry, and cognitive inquiry skills development all appear to have been adequately achieved in Theme I. Some understanding of role responsibilities was achieved, but the criterion level was not met. In view of the large majority of student mean scores reported which met or exceeded their respective criterion level, the null hypothesis is rejected.

Objective 6

Objective 6: To determine whether there are significant differences in student outcomes in biology content knowledge, cognitive inquiry skills and affective qualities of inquiry between students in the following subgroups:

1. Students in classes of IRA teachers using the BSCS Yellow Version text.
2. Students in classes of experienced IRA teachers using the BSCS Blue Version text.
3. Students in classes of non-IRA teachers using the BSCS Yellow Version text.

Hypothesis 6: There is no significant difference in student outcomes--biology content knowledge, cognitive inquiry skills and affective qualities of inquiry as measured by the instruments described in Objective 2--among students grouped by classes of IRA teachers using BSCS Yellow Version, experienced IRA teachers using BSCS Blue Version, or non-IRA teachers using BSCS Yellow Version.

Data Analyses/Results: It is important to first identify which teachers' students were included for these analyses. As DAT scores became available it was readily noticed that the DAT mean scores for students in the three groups given above (IRA Yellow Version classes, IRA Blue Version classes, non-IRA Yellow Version classes) were not equal. Particularly, IRA Yellow Version classes were well below the other student groups. Since it would be inappropriate to simply eliminate selected students with low DAT scores from the analyses, a decision was made to delete groups of students with low DAT mean scores. Thus teacher 01's students (mean score DAT-Verbal = 24.03; mean score DAT-Numerical = 17.73) were deleted as a group. (It should also be noted that teacher 01 did not meet criteria for adequate IRA implementation, and therefore student outcomes would not be considered valid IRA results.) In addition, teacher group 10's students (mean score DAT-Verbal = 28.22; mean score DAT-Numerical = 18.11) were deleted as a group. (Teacher group 10 represented a unique team teaching implementation design with no matching control group on this variable.) These deletions raised the IRA Yellow Version students' mean DAT scores from 29.01 to 30.71 on the Verbal and from 21.12 to 23.87 on the Numerical. This is an increase from approximately the 60th to 65th percentile on the Verbal and from the 40th to the 50th percentile on the Numerical (using 10th grade, first semester norms). Therefore all analyses using IRA Yellow Version scores include data from students of all teachers except teacher 01 and teacher group 10.

Students' scores from all eight control teachers (non-IRA Yellow Version) are included in the EIB and CFE analyses. Three Teachers did not administer the BSBI.

The control group (students of all eight teachers) had a mean DAT-Verbal score of 32.73 (70th percentile on 10th grade first semester norms) and a mean DAT-Numerical score of 26.40 (63rd percentile). These DAT mean scores were not significantly different for the students of the five teachers included in the BSBI analysis.

In order to determine if the primary experimental (IRA-Yellow) group student means for verbal and numerical ability were different from the respective means for the control group, a t-test was utilized. The results are shown in Table 6-1.

TABLE 6-1: Comparison of IRA and non-IRA Yellow Version Students' DAT-Verbal and Numerical Mean Scores.

	DAT - ($\bar{X}$)		S.D.		N		t	P
	IRA	N-IRA	IRA	N-IRA	IRA	N-IRA		
Verbal	30.71	32.73	9.05	8.99	668	487	3.74	.01
Numeric	23.87	26.40	7.28	7.49	656	487	5.62	.01

Thus the control (non-IRA Yellow Version) group had significantly superior DAT-Verbal and Numerical ability over the experimental group (IRA-Yellow Version) used in the following analyses. However, percentile comparisons, as noted earlier, were improved by the deletion of teachers 01 and 10. Further depletion of the experimental group to raise mean DAT scores did not seem warranted since mean DAT scores would not be greatly improved unless a large number of groups were deleted.

The experienced IRA Blue Version teachers reported a student DAT-Verbal mean score of 31.72 (68th percentile) and a student DAT-Numerical mean score of 24.68 (55th percentile). Teacher 64 did not report DAT scores but it is assumed his students are nearly the same since they are within the same district as students of teachers 61 and 62. Note that CFE and BSBI analyses included students from all four of these teachers. EIB analyses, however, include data only from one teacher, 64; the others did not administer the EIB instrument.

In order to determine if there are any significant differences among three groups of teachers' students on any of the posttest scores, a one-way analysis of variance was applied to each of the student outcome variables. The results of these analyses are presented in Tables 6-2 through 6-9.

Note that the EIB subscales reported in previous analyses are not included. Data from non-IRA and experienced IRA teachers was not scored by subscales. The EIB-Part 1A score includes subscales I, III and 12 items in subscale IV. The EIB-Part 1B score includes 12 additional items from subscale IV and subscales V and VI.

TABLE 6-2: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on EIB-1A Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	607	20.71	4.89	20.41	.0000
Non-IRA - Yellow	307	18.33	6.08		
IRA - Blue	29	19.59	5.34		

TABLE 6-3: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on EIB-1B Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	592	41.35	7.73	22.38	.0000
Non-IRA - Yellow	294	37.33	9.95		
IRA - Blue	29	41.48	7.40		

TABLE 6-4: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on BSBI Subscale A (Curiosity) Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	435	2.73	.67	6.49	.0020
Non-IRA - Yellow	141	2.53	.73		
IRA - Blue	107	2.81	.66		

TABLE 6-5: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on BSBI Subscale B (Openness) Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	435	3.74	.67	18.59	.0000
Non-IRA - Yellow	141	3.37	.74		
IRA - Blue	107	3.79	.58		

TABLE 6-6: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on BSBI Subscale C (Satisfaction) Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	435	3.61	.68	5.18	.0061
Non-IRA - Yellow	141	3.46	.75		
IRA - Blue	107	3.74	.68		

TABLE 6-7: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on BSBI Subscale D (Responsibility) Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	435	3.90	1.05	9.07	.0003
Non-IRA - Yellow	141	3.58	1.03		
IRA - Blue	107	4.12	.94		

TABLE 6-8: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on BSBI Total Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	435	13.98	2.28	17.05	.0000
Non-IRA - Yellow	141	12.93	2.53		
IRA - Blue	107	14.46	1.92		

TABLE 6-9: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on CFE Posttest Student Mean Scores

IRA - Yellow	558	21.22	6.25	39.63	.0000
Non-IRA - Yellow	310	24.17	6.39		
IRA - Blue	89	17.97	7.22		

Application of the Hartleys F_{max} test to each analysis demonstrated that the homogeneity of variance assumption underlying analysis of variance was satisfied in each case.

From Tables 6-2 to 6-9 it can be seen that all of the F ratios for comparing the three groups of teachers are significant beyond the .01 level of significance, indicating that there are significant differences among the posttest means for all of the student outcome variables. In order to determine which pairwise means are significantly different, the Newman-Keuls A Posteriori test was computed for all pairs of means. The results of this analysis are presented in Table 6-10.

TABLE 6-10: Table of Pairwise Differences at the .05 Level of Significance as Indicated by the Newman-Keuls A Posteriori Test

TEST	GROUP 1≠ GROUP 2	GROUP 1 GROUP 3	GROUP 2 GROUP 3
EIB 1A	*		
EIB 1B	*		*
BSBI A	*		*
BSBI B	*		*
BSBI C			*
BSBI D	*		*
BSBI Total	*		*
CFE	*	*	*

≠Groups:
1. IRA - Yellow Version
2. Non-IRA - Yellow Version
3. IRA - Blue Version

All of the comparisons of the IRA Yellow Version teachers' students with the non-IRA Yellow Version teachers' students were significant (P = .05) except for the BSBI subscale C score. Of those comparisons showing a significant difference, the IRA Yellow Version teachers' students were significantly higher for all of these differences except for the CFE scores. On the CFE, the non-IRA Yellow Version teachers' students scored significantly higher than both the IRA Yellow Version and IRA Blue Version teachers' students, and the IRA Yellow Version students scored significantly higher than the IRA Blue Version students.

All of the comparisons of the students of the IRA Blue Version teachers with the non-IRA Yellow Version students were significant (P = .05) except for the EIB 1A scores. For those comparisons showing a significant difference, the students of the IRA Blue Version teachers were significantly higher than the non-IRA students in all comparisons except for the CFE scores. As noted above, the IRA Blue Version students were significantly below both the IRA and the non-IRA Yellow Version students on the CFE.

The only pairwise comparison between the IRA Yellow Version with the IRA Blue Version students that was significant was on the CFE. All of the other comparisons involving these two groups of students were not significant at the .05 level.

Interpretation: Despite the superiority by the control group in verbal and numerical ability as measured by the DAT, the IRA student groups had significantly superior posttest scores to the control group in cognitive inquiry and affective qualities of inquiry. These results are particularly meaningful for evaluating the effectiveness of the IRA program in light of the fact that the IRA program has been developed to operationalize the attitudinal and cognitive inquiry objectives delineated in Inquiry Objectives in the Teaching of Biology (this basis is discussed in the CONTEXT section of this report). These results indicate that the IRA program is an effective teaching approach for developing cognitive inquiry skills and affective qualities of inquiry which have been previously recognized by science educators as important goals of science teaching.

Note that these results on the EIB and BSBI analyses also support the validity of the IRA Yellow Version students' pre to post gains presented and discussed in Objective 3A.

With respect to the posttest biology content instrument, CFE, student mean scores for the non-IRA-Yellow Version group significantly exceeded the scores for the IRA-Yellow Version group. This finding should be interpreted in terms of the differences in the two student groups on DAT scores (Verbal and Numerical), the standard error of measurement reported in the CFE Manual and the quantity of content coverage in the IRA Yellow Version Groups.

Part of this difference may be due to the significant differences in the DAT scores (Verbal and Numerical) reported on Table 6-1 which was significantly higher for the non-IRA Yellow Version group.

Another factor to consider is that the difference in the mean scores for the two groups (2.95) is within the standard error of measurement (3.1 to 3.3) reported in the CFE Manual.

It is possible that some of the difference in the obtained scores can be attributed to measurement error and does not represent "true" difference in the scores of the two groups.

Note, that the first two IRA themes treat 41 percent of the chapters in the BSBS Yellow Version text; the majority of IRA Yellow Version teachers completed only 11 percent of Theme III activities. The low extent of biology content treatment indicated by this information, plus IRA teachers own statements that content treatment was reduced from previous years when IRA was not used, indicate that the lower CFE scores may be due in part to reduced biology content treatment. (Interviews of both IRA teachers and non-IRA teachers in previous IRA studies showed that IRA teachers treated at least 25 percent fewer text chapters than non-IRA; it is reasonable to assume that this disparity of treatment also existed in the 1972-73 field test study.)

In light of the probable disparity of content treatment and differences in CFE posttest scores, it can be implied that in using the IRA program and in thereby expanding course objectives to include cognitive inquiry and affective qualities development, teachers must be aware that some reduction in the scope of biology content treated may be necessary. It should be pointed out, however, that in previous studies (1969-70, 1971-72) IRA classes scored significantly higher on CFE posttests than non-equivalent non-IRA classes, groups with equivalent DAT scores were used in these studies.

The Yellow Version IRA groups scored significantly higher than Blue Version IRA groups on CFE scores. There appears to be no particular reason to believe that differences in DAT scores, measuring error, or differences in the treatment of subject matter coverage in the course should account for these differences. Also previous experience in studies conducted in local IRA Blue Version classes have shown that the students scored much higher than found in this study.

Part of the difference can probably be attributed to fifty percent of the students included in the Blue Version sample for CFE being 9th graders. Based on previous experience with 9th grade students the investigators as well as the CFE Manual authors have found considerable differences in scores favoring 10th graders. Otherwise, the differences in these results remain unexplained.

In summary, the students of IRA Yellow Version teachers have shown significantly higher posttest scores on instruments measuring cognitive inquiry skills and affective qualities of inquiry than students of non-IRA Yellow Version teachers. This suggests that the IRA program is an effective teaching methodology for the development of cognitive inquiry and affective qualities of inquiry. Students of non-IRA Yellow Version teachers have shown significantly higher posttest scores on an instrument measuring biology content knowledge than students of IRA Yellow Version teachers. This difference may be due in part to non-equivelent verbal and numerical abilities of the IRA and non-IRA students, error in measurement and to the probable disparity in biology content treated in the IRA and non-IRA classes. This result is also not consistent with results of two previous studies.

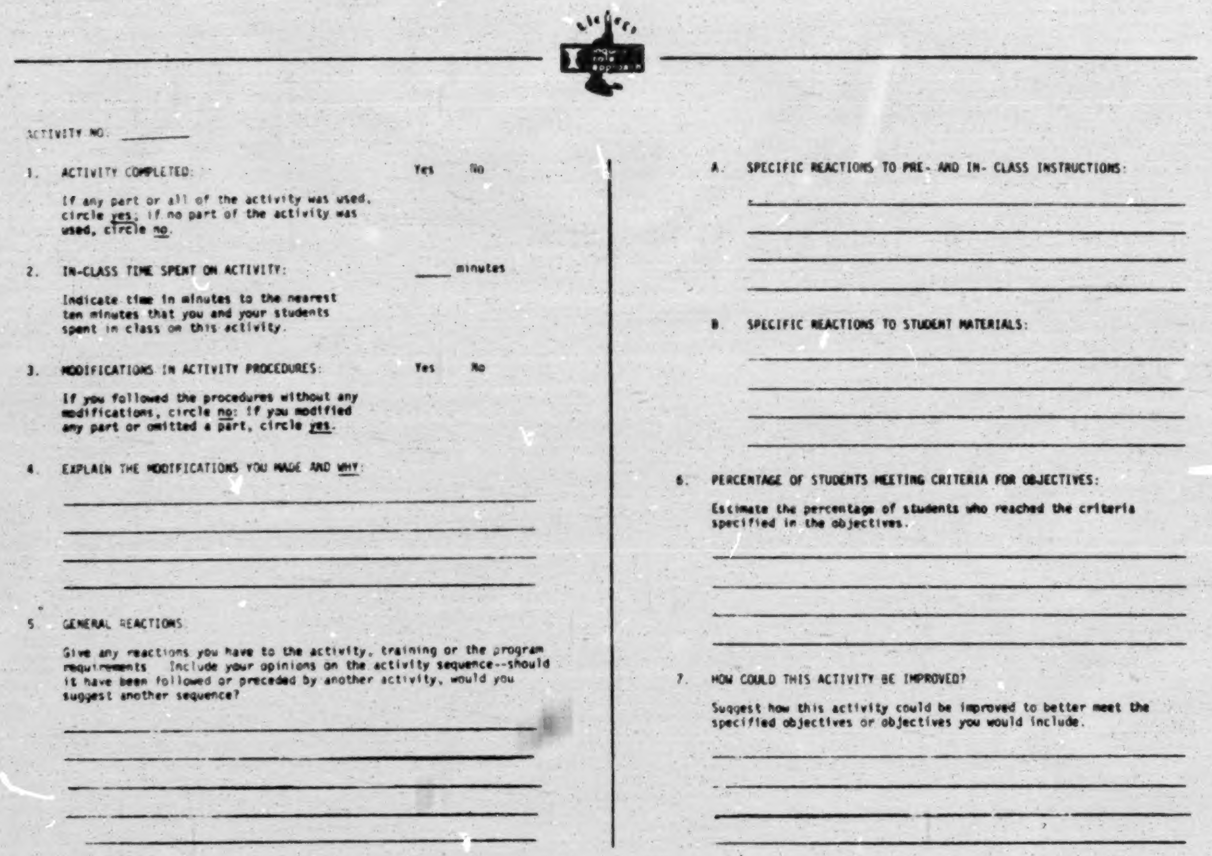
The Yellow Version IRA classes have shown significantly higher posttest scores on the CFE instrument than the Blue Version IRA classes. Other than the grade level difference in the two groups the results appear inconsistent with past studies.

Objective 7

Objective 7 - To determine what revisions in the program materials are indicated by the teacher responses.

Hypotheses 7 - None; this is a descriptive objective. Information for the revision recommendations will be taken primarily from sections 3, 4, 5 and 7 of the Teacher's Log and secondarily from other records of teacher feedback (reports from on-site visits, memoranda and letters from teachers, notes regarding telephone or personal communication with teachers, etc.)

Data Analysis/Results: All field test teachers were asked to complete a Teacher's Log form following performance of each Inquiry Role Approach activity.



ACTIVITY NO: ________

1. ACTIVITY COMPLETED: Yes No

If any part or all of the activity was used, circle yes; if no part of the activity was used, circle no.

2. IN-CLASS TIME SPENT ON ACTIVITY: ____ minutes

Indicate time in minutes to the nearest ten minutes that you and your students spent in class on this activity.

3. MODIFICATIONS IN ACTIVITY PROCEDURES: Yes No

If you followed the procedures without any modifications, circle no; if you modified any part or omitted a part, circle yes.

4. EXPLAIN THE MODIFICATIONS YOU MADE AND WHY:

5. GENERAL REACTIONS

Give any reactions you have to the activity, training or the program requirements. Include your opinions on the activity sequence--should it have been followed or preceded by another activity, would you suggest another sequence?

A. SPECIFIC REACTIONS TO PRE- AND IN- CLASS INSTRUCTIONS:

B. SPECIFIC REACTIONS TO STUDENT MATERIALS:

6. PERCENTAGE OF STUDENTS MEETING CRITERIA FOR OBJECTIVES:

Estimate the percentage of students who reached the criteria specified in the objectives.

7. HOW COULD THIS ACTIVITY BE IMPROVED?

Suggest how this activity could be improved to better meet the specified objectives or objectives you would include.

Information contained in the Teacher's Log report enabled the field test staff to:

a. Sense the degree of implementation enroute during the field test and to determine occasional need for special onsite monitoring and for other contact with IRA teachers or trainers.

b. Compile a record of teacher and student performance on an activity-by-activity basis for later summary.

Upon completion of the field test the log data were summarized and this information provided input for decisions on deletion, revision, or other modification of Inquiry Role Approach activities. The summary process was as follows: Each Teacher's Log was read and comments relative to Objective 7 of the field test were recorded on a blank Teacher's Log, referred to as data log. An Activity Summary Form was then developed to record summaries of the information from the data logs.

ACTIVITY SUMMARY FORM

Activity No. ____

Percent of teachers who did the activity in whole or part ____.

Percent of teachers performing the activity who reported students met criteria ____.

Average percent of students meeting objectives based on teacher logs or from other communications ____.

Summary of feedback from teacher logs or other communications.

Modifications made and why:

General Reactions:

Specific Reactions to Pre- and In-Class Instructions:

Specific Reactions to Student Materials:

- over -

ACTIVITY SUMMARY FORM, p. 2

Status of Activity　　　　Activity No. ____

Leave as is ____

How could activity be improved?

Pre- and In-class instructions ____________________

Student Materials ____________________

Change sequence ____________________

Split into two or more activities ____________________

Combine with other activity(ies) ____________________

Four field test staff were involved in the summary process, two staff members summarized data logs from all activities, two other staff members summarized data logs from separate groups of activities. The result was that all data logs were summarized by at least three staff members. The output was then combined into two charts: Activity Summary Chart A and Activity Summary Chart B.

In reviewing Activity Summary Chart A the reader should note that column 2, "Number of Activity-Performing Teachers Who Reported Percent of Students Meeting Criteria" is usually different from the total number of teachers reporting in column 1. Therefore, the figures in column 3 are based on varying population sizes dependent on the number of teachers who actually reported student criterion performance in their log reports to McREL. A draft of the second chart "Activity Summary Chart B" was prepared based on activities summaries (as described above) and circulated back to the four staff members who prepared the summaries for review to determine if the decisions noted for each activity number accurately reflected their opinion of the meaning of the teacher logs. A conference was then attended by all staff members involved in the log summary process to finalize Activity Summary Chart B and agree on the specific nature of recommendations by a review of teacher participant logs.

During the summarizing process, the field test staff occasionally encountered a collection of teacher log responses on a given activity ranging from "very specific and clear as usual" to "students needed a lot of explanation..." in the general reactions section. (These quotes are taken from Activity 117 log, category 5: general reactions.) In the event of wide variance of this sort, the investigators used as primary input to determining the overall quality of an activity, the figures as recorded in Activity Summary Chart A. It was significant that in such situations a wide variance of teacher reaction was usually associated with entries of 6 and above for column 2 and 7 and above for column 3. The decision was reached to leave such activities substantially unchanged or else provide for minor changes for clarification. However, major changes of substance, sequence changes, or combinations were not recommended in such cases.

Theme III Activities: It is important to note that no previously inexperienced IRA teachers reported data on Theme III. The previously experienced teachers did report on use of Theme III. This theme is very "open" in its provision of optional activities and a completely unstructured LEIB. Due, therefore, to diminishing log returns for Theme III only Chart A figures are given. The data base is not sufficient to generate specific recommendations for revision, hence entries are not made in Chart B for this theme.

Interpretation: Of the thirty six activities included in Activity Summary Chart B, fourteen require major changes. It should be noted, however, that changes generally dealt with better directions to the teacher (more direction to execute activity, more accurate time estimates, more complete discussion of expected student outcomes or assessment of outcomes, etc.) or changes in clarity or usefulness in student materials (shorten student forms, clarify statements, etc.). Recommendations to delete activities or major parts of activities, to redirect the activity to new goals, to substitute other activities, etc., were only suggested in response to the introductory activities, 101 to 105. Even when such changes were suggested, common elements of an initial orientation to the IRA program were found in all teacher suggested revisions.

Teacher feedback suggests that the IRA materials were found to be adequate for implementation in the classroom and generally satisfactory to teachers in terms of usability.

ACTIVITY SUMMARY CHART A

Activity Number and Title	Number of Teachers Performing Activity in Whole or Part	Number of Activity-Performing Teachers Who Reported Percent of Students Meeting Criteria	$\overline{X}$ % of Students Meeting Criteria Based on Teacher Logs or Other Communication
000 General Section	*XXX	XXX	XXX
101 Pre-Testing	11	XXX	XXX
102 Introduction to Biological Inquiry	10	10	86.44
103 Inquiry Role Approach Dimensions	10	6	78
104 Lost on the Moon	11	10	83.11
105 Student Feedback Summary	11	9	88.50
106 Life in Unexpected Places	10	7	73.50
107 Inquiry Process	11	9	63.88
108 The Compound Microscope - A Scientific Tool	10	7	86.83
109 Life From Nonlife (Inquiry 2-1)	9	7	74.16
110 Introduction - Biological Succession (Inquiry (Inquiry 36-1)	10	7	74.16

*No entry required

ACTIVITY SUMMARY CHART A

Activity Number and Title	Number of Teachers Performing Activity in Whole or Part	Number of Activity-Performing Teachers Who Reported Percent of Students Meeting Criteria	$\bar{X}$ % of Students Meeting Criteria Based on Teacher Logs or Other Communication
111 Sampling Populations (Invitation #9)	10	8	85.71
112 Estimating Relative Population Density	9	8	80
113 Biological Succession (Inquiry 36-1)	11	5	70.75
114 Checks & Balances in Nature	10	8	66.42
115 Student Feedback - Roles & Social Process	11	8	82.86
116 Introduction to LEIB 1	11	7	75.83
117 Random & Systematic Error	11	9	76.63
118 Laboratory Exploration in Biology	11	*XXX	XXX
119 Class Presentation of LEIB 1	10	5	84
120 Replanning LEIB 1	9	5	82
121 Theme I Assessment	11	2	78.5
122 Testing on Class Procedures	11	XXX	XXX

*No entry required

ACTIVITY SUMMARY CHART A

tivity Number and Title	Number of Teachers Performing Activity in Whole or Part	Number of Activity-Performing Teachers Who Reported Percent of Students Meeting Criteria	$\overline{X}$ % of Students Meeting Criteria Based on Teacher Logs or Other Communication
0 General Section	*XXX	XXX	XXX
1 Setting Goals for Theme II	9	4	80
2 Reading Selections: Priestley & Van Helmont	10	6	84.16
3 Introduction to IEIB 2	10	6	76.5
4 Photosynthesis & Respiration	10	6	70.5
5 Understanding Basic Chemistry	10	6	71.66
6 Living Chemistry	10	7	79.42
7 Physiology of Cells	10	6	85.16
8 The Living Chemistry of Cells	10	7	78.29
9 Plants and Photosynthesis	9	4	80.75
0 Chemistry in Cells of Photosynthetic Plants	9	5	80.80
1 Laboratory Explorations in Biology 2	10	XXX	XXX

*No entry required

ACTIVITY SUMMARY CHART A

Activity Number and Title	Number of Teachers Performing Activity in Whole or Part	Number of Activity-Performing Teachers Who Reported Percent of Students Meeting Criteria	$\overline{X}$ % of Students Meeting Criteria Based on Teacher Logs or Other Communication
212 Class Presentation of LEIB 2	8	3	90
213 Replanning LEIB 2	3	2	86
214 Theme II Assessment	5	1	75
215 Testing on Class Procedures	*XXX	XXX	XXX
000 General	XXX	XXX	XXX
301 Setting Goals For Theme III	2	2	70
302 Introduction to LEIB 3	2	2	77.5
303 Suggested Pre-LEIB Activities	2	2	81.50
304 Inquiry Into Inquiry	1	1	60
305 Laboratory Explorations in Biology	1	XXX	XXX
306 Class Presentation of LEIB 3	1		
307 Replanning LEIB 3	1		
308 Theme III Assessment	1		
309 Post-Testing	11		

*No entry required

THEME I

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
101			1. Spread testing over several weeks with knowledge that BSBI and V&P-C are most affected by delay.		These comments apply to Activities 101 thru 104: These activities may be varied in sequence provided that 101 testing comes first (certain tests) and is then intermingled. Either 102 is eliminated and/or 103 shortened, eliminated, or delayed so as to be the fourth activity. These changes were made by individual and groups of "successfully implementing teachers."
102			1. Indicate other film-loops that can be substituted that have comparable value to reach same objective and give guidelines to use substitute. 2. An additional period may be needed.		If Pre tests are administered over several activities the limitations will be noted.
103			1. This activity modified in different ways in field test. Therefore it can be optional or subject to teacher modification.		

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
104			1. Rewrite "Lost on Moon" to some topic nearer to biology and develop introduction to roles in discussion following "Lost on..." 2. Return to original idea of having entire class go thru problems as individual first, then as a group. 3. Revise assessment form 104-5 accordingly. 4. Form 104-7 is revised in accordance with #2 above.	Combine assessment form with 105-3.	
105			1. Revise 105-3 as necessitated by combination.	Combine with 104.	
106			1. Link more closely with lab-text material if "old lab" is used. 2. Choose a simpler lab exercise. Administer role quiz following laboratory experience rather than prior to it.		

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
107			Simplify the inquiry guide in the following ways: 1. Word the interpretations so they state the concept and show direction, i.e., they are false when the statement is false, etc. The interpretation should not necessarily be in question form. 2. In statement B, insert word "only." 3. In statement C consider, "The students could accept the outside temperature as facts but it should not be considered as data." Interpretation: "All facts are not considered to be data." 4. Check textbook on use of plasmodium. Change interpretation of "D" in 107-2 to statement,		

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
107 (cont)			4. (cont) "An inquiry process is completed by forming a complete hypothesis--one which includes both an explanation of known facts and a prediction." 5. Take inference E and application question out of student material and put it in Teacher's Manual for "powerful students." An option might be to offer it for group work. 6. Materials 107-4 - needs to be used once, 107-7 could be eliminated. 7. On 107-9 -- No. 2 (drop point No. 3). 8. On 107-9 -- No. 2, reword present point No. 4 to: "Organize these key words into your statement presenting the same basic issue presented in inquiry guide statement."		

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
107 (cont)			9. On 107-9 -- Reword No. 4 as follows: "Search for information using the references given to support or refute the issue in the interpretation." 10. On 107-9 -- No. 5 to be reworded.		
108	X				
109		1. Put in teacher's materials instructions to prepare second set of cultures so there is not attrition from long delay from first laboratory.			
110		1. Objectives possibly need to be rewritten for students. Teachers not having access to hemocytometers should have option suggested of using high power estimation with directions for using this procedure in an			

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
110 (cont)		1. (cont) optimum way. Otherwise, keep in present hemocytometer material.			
111		1. Clarify the directions to allow teachers to follow various lines of discussion as suggested by students.			
112		1. Advise teachers to prepare new cultures each day. Instructions be given to avoid expecting too much perfection on activities. Move along--check microbiology (lab block) for sameness with 112-3 & 112-4.	1. Need some instructions for teachers on care and use of hemocytometers. Teacher should preteach the TA's on use of hemocytometer and also give TA's a sample division of work for team members.		

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
112 (cont)		2. Insert a paragraph on preparation of yeast cultures.			
113			1. Retain this type of activity but substitute another lab that is not so difficult in terms of equipment required.		
114			1. Inquiry Guide should be condensed to 5 statements. All application questions should be discussed in teamwork. The cognitive assessment response key should identify certain responses where T and F are equally appropriate.		*On-site visits indicated students (and teachers) felt too much paper in Theme I was concerned with roles; students did not know which role sheets to follow since "new" sheets were continually being issued. This suggests a review of role development in 106, 107, 114, and 118 and an attempt to reduce and clarify papers related to roles. 114 role work is particularly long and involved and (except for special problem sheets) appears to duplicate Activity 107 papers. *These comments are based on on-site visit records and/or other feedback available at McREL.

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
115		1. Reword No. 8 on 115-3. 2. 115-1 - Suggested key assessment at bottom of p. 2. Should read "responses from hypothetical situation" strike "suggested key" wording. 3. Provide directions to teacher to use responses from other classes as a model for classes not doing well.	1. Reduce length of Form 115-3.		
116	X				
117	X				
118		1. Give realistic time expectation.			*Teachers indicated that the 118 role sheet (which compared each role's responsibilities in lab, guide work, and LEIB work) was confusing and therefore not useful to students. This is consistent with overall reaction to role work as expressed in comment for Activity 114. More teacher directions are necessary to give teachers

*These comments are based on on-site visit records and/or other feedback available at McREL.

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
118 (cont)					(cont) better understanding that they are free to intervene as necessary if students are not able to draw major ideas out of class reporting.
119	X	1. Shorten 119-11.			
120	X				
121		1. Use correct items as indicated (Teacher 03).	1. Shorten test sequence.		
122		1. Give CAQ (recommend) and V&P other than at "testing time."			

THEME II

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
201		1. Reduce complexity by separate team formation and Theme II introduction sections.			
202			1. Suggest topics for teachers on which they may be able to formulate questions. Time requirements should be revised.		
203	X				
204		1. Directions to teachers should include some suggested procedures for teacher's role in presenting content. Procedures might include reading the objectives to students, lectures, assigning homework, using overhead transparencies and evaluation.			

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
205		1. Refer teachers to instructions on how to write inquiry guides which appears early in Theme III for getting ideas.			
206		1. Teacher should request other students to take notes on reports.	1. Give instructions to teachers to work with each presenting team to: a. Review content. b. Review method of presentation. c. Suggest indirectly ways of improving presentation.		
207			1. Give specific guidelines for scoring oral and written reports. 2. Revise Form 009 and 0010 accordingly and grade it.		
208		1. Give teachers directions to distribute duplicated lab results.	1. Correct errors in 208-3 and 208-4. As noted in logs.		

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
209			1. Revisions here similar to those recommended for 206 and 207. 2. Student objectives 2 and 3 should be changed to be consistent with changes on 009 and 0010.		
210		1. Need instructions to teachers to indicate that inasmuch as this is the last inquiry guide before the LEIB, they should feel free to use interest-generating techniques (student use of overhead on board, freedom to use reference(s) other than text) to counter some tendencies toward "inquiry guide fatigue." This inquiry guide is important as a transition to the LEIB.			

ACTIVITY SUMMARY CHART B

Recommendations for activity revision based on interpretation of teacher logs.

Activity Number	Substantially Unchanged	Minor Changes for Clarification	Major Changes of Substance	Combine	Comments
211		1. Teachers should be instructed to help students create their own problems as well as use the list of problems compiled in previous activities.			
212	X				
213	(insufficient data)				
214	X				
215		1. Give CAQ and V&P other than at "testing time."			

Ex Post Facto Studies

Supplementary Objective 1

Supplementary Objective 1: To compare the percent of IRA classes emphasizing the CAQ factors application, analysis, synthesis, evaluation, discussion, independence, and divergence, to the percent of average and gifted classes emphasizing these factors.

Supplementary Hypothesis 1: None; this is a descriptive objective.

Data Analysis/Results: Steele[8] has reported studies which compare the percent of gifted and average classes emphasizing factors identified on the CAQ. Seven of the nine factors utilized for evaluation of IRA implementation in the field test were included in Steele's report. The percent of average and gifted classes emphasizing these seven factors, and the percent of IRA classes emphasizing these factors, are given in Table S1-1.

TABLE S1-1: Patterns of Instructional Climate of Average and Gifted Classes (as reported by Steele) and of Inquiry Role Approach Classes

CAQ DIMENSIONS	FACTORS	PERCENT OF CLASSES EMPHASIZING FACTOR AVERAGE (N=69)	 GIFTED (N=62)	 IRA (N=8)*
Higher Thought Processes	Application	--	52%	100%
	Analysis	58%	81%	100%
	Synthesis	--	40%	62%
	Evaluation	--	31%	50%
Classroom Focus	Discussion	30%	89%	62%
Classroom Climate	Independence	28%	76%	85%
	Divergence	69%	97%	100%

*IRA teachers included are those who administered CAQ at the end of Theme II as reported on Table 1-1 (see section on Objective 1).

As seen in Table S1-1, the percent of IRA classes emphasizing each factor surpassed the percent of both average and gifted classes for all factors except Discussion. The percent of IRA classes emphasizing Discussion did, however, well exceed the percent of average classes emphasizing this factor.

The data here represents results in classes of 12 IRA teachers. (Teachers 10 through 14 were, of course, treated as one entry; therefore N = 8 on the table.) Three teachers did not administer the CAQ at the end of Theme II.

Interpretation: In terms of patterns of instructional climate which have been identified as important elements in an IRA class setting, IRA classes performed well when compared to previously reported results for average and gifted classes.

Supplementary Objective 2

Supplementary Objective 2: To determine whether there are significant differences in student outcomes in biology content knowledge, cognitive inquiry skills and affective qualities of inquiry between students of IRA Yellow Version teachers with previous IRA experience and students of IRA Yellow Version teachers without previous IRA experience.

Supplementary Hypothesis 2: There are no significant differences in student outcomes--biology content knowledge, cognitive inquiry skills and affective qualities of inquiry as measured by the instruments described in Objective 2--between students of IRA Yellow Version teachers with previous IRA experience and students of IRA Yellow Version teachers without previous IRA experience.

Data Analyses/Results: In order to test Supplementary Hypothesis 2, a repeated measures analysis of variance was run on each of the 11 student outcome variables. The results of these analyses are presented in Tables S2-1 through S2-11. Note that the group identified as "First Year" represents students of IRA Yellow Version teachers without previous IRA experience (teachers 01, 20, 21, 22, 30, 31, 40 and teacher group 10). The group identified as "Experienced" represents students of IRA Yellow Version teachers with previous IRA experience (teachers 02, 03, 04).

TABLE S2-1: Comparison of Mean Scores of Students of First Year and Experienced IRA Teachers for EIB III (Formulate a Hypothesis)

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	266	10.21	11.53	10.87
Experienced	89	10.62	12.26	11.44
Test Means	355	10.31	11.71	

F Groups = 3.06, P = .08; F Tests = 41.78, P = .00, F GxT = .41, P = .53

TABLE S2-2: Comparison of Mean Scores of Students of First [illegible] and Experienced IRA Teachers for EIB IV (Design a Stu[illegible])

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	214	10.40	17.64	14.02
Experienced	67	16.25	17.93	17.09
Test Means	281	11.79	17.71	

F Groups = 67.9, P = .00; F Tests = 589.4, P = .00, F GxT = 95.0, P = .00

TABLE S2-3: Comparison of Mean Scores of Students of First Year and Experienced IRA Teachers for EIB V (Interpretation of Data)

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	276	9.08	23.34	16.21
Experienced	80	20.46	25.11	22.79
Test Means	356	11.63	23.74	

F Groups = 139.45, P = .00; F Tests = 1003, P = .00, F Gx[illegible] = 110, P = .00

TABLE S2-4: Comparison of Mean Scores of Students of First Year and Experienced IRA Teachers for EIB VI (Synthesizing Knowledge)

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	259	5.18	6.95	6.07
Experienced	75	6.01	7.19	6.60
Test Means	334	5.37	7.01	

F Groups = 7.81, P = .01; F Tests = 120.1, P = .00, F GxT = 2.8, P = .09

TABLE S2-5: Comparison of Mean Scores of Students of First Year and Experienced IRA Teachers for EIB Total Scores

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	214	35.95	60.82	48.38
Experienced	67	53.87	63.97	58.92
Test Means	281	40.22	61.57	

F Groups = 71.09, P = .00; F Tests = 832, P = .00, F GxT = 72.3, P = .00

TABLE S2-6: Comparison of Mean Scores of Students of First Year and Experienced IRA Teachers for BSBI Subscale A (Curiosity)

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	248	2.64	2.75	2.69
Experienced	68	2.46	2.53	2.50
Test Means	336	2.60	2.70	

F Groups = 7.11, P = .01; F Tests = 5.44, P = .02, F GxT = .7, P = .68

TABLE S2-7: Comparison of Mean Scores of Students of First Year and Experienced IRA Teachers for BSBI Subscale B (Openness)

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	248	3.58	3.65	3.61
Experienced	68	3.32	3.63	3.48
Test Means	336	3.52	3.64	

F Groups = 2.19, P = .14; F Tests = 8.12, P = .01, F GxT = 5.24, P = .02

TABLE S2-8: Comparison of Mean Scores of Students of First Year and Experienced IRA Teachers for BSBI Subscale C (Satisfaction)

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	248	3.66	3.57	3.62
Experienced	68	3.60	3.47	3.53
Test Means	336	3.65	3.55	

F Groups = 1.19, P = .28; F Tests = 7.45, P = .01, F GxT = .21, P = .66

TABLE S2-9: Comparison of Mean Scores of Students of First Year and Experienced IRA Teachers for BSBI Subscale D (Responsibility)

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	248	3.62	3.73	3.67
Experienced	68	3.47	3.52	3.49
Test Means	336	3.58	3.68	

F Groups = 7.11, P = .01; F Tests = 2.31, P = .13, F GxT = .18, P = .67

TABLE S2-10: Comparison of Mean Scores of Students of First Year and Experienced IRA Teachers for BSBI Total

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	248	13.50	13.69	13.59
Experienced	68	12.86	13.14	13.00
Test Means	316	13.36	13.57	

F Groups = 4.11, P = .04; F Tests = 2.90, P = .09, F GxT = .08, P = .77

TABLE S2-11: Comparison of Mean Scores of Students of First Year and Experienced IRA Teachers for CFE

GROUP	N	PRETEST MEAN SCORE	POSTTEST MEAN SCORE	GROUP MEAN SCORE
First Year	276	18.17	19.39	18.78
Experienced	73	17.60	18.67	18.14
Test Means	349	18.05	19.25	

F Groups = 1.06, P = .30; F Tests = 12.45, P = .00, F GxT = .04, P = .84

The means and F ratios for comparing the EIB III scores for the first year versus the experienced teachers are presented in Table S2-1. As indicated in the table, only the F ratio for comparing the pretest means with the posttest means is significant. Thus, for both groups of teachers combined, there was a significant increase in the EIB III scores. Both groups improved relatively the same amount since the interaction effect was not significant.

The results for the EIB IV scores are presented in Table S2-2. As indicated in the table, all of the F ratios are significant beyond the .01 level. Thus there is a significant difference between the two groups of teachers as well as a significant difference between pretest and posttest scores. The interaction is also significant indicating that the gain from pretest to posttest was not similar for both groups of teachers. As Kirk[9] points out, whenever there is a significant interaction effect, interpretation of the tests of the main effects must be qualified. From an inspection of Table S2-2, it can be noted that there is a six point difference between pretest and posttest means as well as a three point difference between the means for the two groups of teachers. It also can be noted that the gain for the first year teachers was around seven points whereas the gain for the experienced teachers was only about a point and a half. As indicated by the significant interaction, this difference in gain is a significant difference.

The same results were found for the EIB V scores. All of the F ratios are significant, and as can be noted in Table S2-3 the gain from pretest to posttest for the first year teachers was much greater than for the experienced teachers. As with the EIB IV scores, the pretest scores for the experienced teachers were higher than for the first year teachers. However, since the first year teachers' students gained significantly more than the experienced teachers' students, the posttest scores are nearly equivalent.

For the EIB VI scores (Table S2-4) both the F ratios for comparing groups and for comparing tests were significant. Again, the experienced teachers' students scored higher than the first year teachers' students on both the pretest and the posttest. Since the interaction effect was not significant, the gains of the two groups of teachers is equivalent. The gain for the first year teachers is about one and a half points whereas for the experienced teachers it is a little over one point.

For the EIB total scores (Table S2-5) the same pattern is evident as with the Part IV and V scores. All of the F ratios are significant beyond the .01 level. The gain from pretest to posttest is significant and the 10 point difference between the first year and experienced teachers' students' scores is also significant, the experienced teachers' students' scores being higher than the scores for the first year teachers. The gains for the two groups were significantly different as indicated by the significant F ratio for the interaction effect. The gain for the first year teachers' students was nearly 25 points whereas the gain for the experienced teachers' students was only around 10 points.

The results for the BSBI scores (Tables S2-6 through S2-10) are somewhat different than the results for the EIB scores. For the BSBI scores, the scores for the first year teachers' students tend to be higher than the scores for the experienced teachers' students. For the BSBI A, D, and total scores, the scores for the first year teachers' students are significantly higher than the mean scores for the experienced teachers' students. None of the interactions for these three BSBI scores are significant indicating that the gain from pretest to posttest is similar for both groups of students. There is a significant difference between pretest and posttest scores for BSBI subscales A, B, and C. Thus, the gains from pretest to posttest for the three subscales A, B, and C are significant. The only F ratio for interaction effects which is significant is for the BSBI subscale B. From Table S2-7 it can be noted that the gain for the first year teachers' students was .07 whereas the gain for the experienced teachers' students was .31.

For the CFE scores (Table S2-11) only the F ratio for comparing pretests with posttests is significant. Thus it appears that both the scores and the gains for the two groups of teachers was similar. However, the gain from pretest to posttest for both groups of students is significant.

Interpretation: While experienced IRA Yellow Version teachers' students demonstrate significantly higher group mean scores over first year IRA Yellow Version teachers' students for EIB IV, V, VI and EIB total, the interaction effects were also significant for EIB IV, V and EIB total. In these three cases, students of first year IRA teachers showed significantly higher pre to post gains. Thus the higher group mean scores of the students of experienced teachers do not allow for any implications regarding IRA experience as related to student cognitive inquiry outcomes.

No pattern emerges from the BSBI analyses except to say that the students of first year IRA teachers generally show higher scores (but not significantly higher in most cases). The CFE results also do not show meaningful differences.

These analyses suggest that success with the IRA program in terms of student outcomes as measured by the EIB, BSBI and CFE is not strongly related to the teachers' extent of IRA experience.

It should be noted that the pre to post comparisons for the combined "first year" plus "experienced" means showed significant gains for 9 of the 11 variables. However, these have not been discussed here since such gains have already been discussed under Objective 3A.

REFERENCES

1. Kruskal, W. and Wallis, W. "Use of Ranks in One-Criterion Variance Analysis," Journal of the American Statistical Association, Vol. 47, 1952, pp. 583-621.

2. Campbell, D. T. and Stanley, J. C. Experimental and Quasi-Experimental Designs for Research, Rand McNally & Co., Chicago, 1963.

3. Koos, E. M. "Technical Report No. 1. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, Explorations in Biology," Mid-continent Regional Educational Laboratory, 1970.

4. Biological Sciences Curriculum Study. Manual for the Comprehensive Final Examination in First-Year Biology, The Psychological Corporation, New York, N. Y., 1966.

5. Biological Sciences Curriculum Study. Quarterly Achievement Tests for BSCS Yellow Version, The Psychological Corporation, New York, N. Y., 1966.

6. Biological Sciences Curriculum Study. Resource Book of Test Items for Biological Science An Inquiry Into Life Second Edition, Educational Programs Improvement Corporation, Boulder, Colo., 1971.

7. Koos, E. M. "Technical Report No. 1. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, Explorations in Biology," Mid-continent Regional Educational Laboratory, 1970.

   Koos, E. M. "Technical Report No. 2. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, Explorations in Biology," Mid-continent Regional Educational Laboratory, 1971.

   Koos, E. M. and Chan, J. Y. "Technical Report No. 3. Criterion-References Tests in Biology," paper presented at American Educational Research Association Meeting, April, 1972.

   Koos, E. M. "Technical Report No. 4. A Report on Developmental Studies of a Series of Measures of Inquiry Skill in Biology, Explorations in Biology," Mid-continent Regional Educational Laboratory, 1972.

8. Steele, J. M., House, E. R., and Kerins, T. "An Instrument for Assessing Instructional Climate Through Low Inference Student Judgments," American Educational Research Journal, 1971, Vol. 8, No. 3, pp. 446-447.

9. Kirk, Roger E. "Experimental Design Procedures for the Behavioral Sciences," Brooks-Cole Publishing Company, Belmont, Calif., 1968, p. 177.

APPENDIX

The appendix for this rep[illegible] contains copies of the following:

1. Teacher's Log
2. Views & Preferences - C
3. Class Activities Questionnaire
4. Comprehensive Final Examination
5. Biology Student Behavior Inventory
6. Explorations in Biology 1A, B, 2A, and B
7. Form 121-4, Social Skills Checklist
8. Form 121-5, Attitude Checklist
9. Form 121-3, Understanding Role Responsibilities Quiz
10. Form 35c, Differentiation Between Roles
11. Form 35d, Differentiation Between Roles
12. Form 21, Assessment of Role Functions
13. Form 26, Assessment of Role Functions
14. Form 121-7, Theme I Assessment Summary Sheet
15. Differential Aptitude Test
16. Letter and questionaire sent to prospective field test participants
17. Replanned Workshop Schedule

Copies of the appendix can be obtained by writing to

Dr. Paul G. Koutnik
Coordinator of Inquiry Skill Research, Development and Adaptation
Mid-continent Regional Educational Laboratory
104 E. Independence Ave.
Kansas City, Missouri 64106